

השפעת ה-AI על שוק העבודה

ויזואליזציה של מידע

דו"ח מסכם

מגישים:

נועה שגב - 210025631

חן מרקוביץ - 211389374

סתיו עוזרי - 206360067

ניב דולב - 323817866

תאריך הגשה:

23.01.2026

קישור לאפליקצית הדאשבורד:

<https://data-visualization---projectapp-xdv3vriyf7wojgvtaqlwva.streamlit.app>

קישור ל-GITHUB

<https://github.com/NivDol/Data-Visualization---ProjectStreamlit>

קישור למקורות מידע

<https://www.kaggle.com/datasets/bismasajjad/global-ai-job-market-and-salary-trends-2025?resource=download>

חלק ראשון - מבוא:

הפרויקט שלנו עוסק בויזואליזציה של נתונים אודות השפעת ה-AI על שוק העבודה העולמי. בשנים האחרונות הפך ה-AI לאחד הגורמים המשמעותיים ביותר המשפיעים על שוק העבודה העולמי.

שילובו של ה-AI בתהליכי עבודה ובקבלת החלטות, שינה באופן רחב את מבנה המשרות, ואת הציפיות מהעובדים.

מטרת העבודה היא לבחון בצורה ויזואלית ומבוססת נתונים את **השפעת ה-AI על שוק העבודה**, וזו השאלה

המרכזית עליה אנחנו רוצים לענות.

לטובת עבודה זו נעשה שימוש בדאטה שנלקחה מאתר Kaggle :

Global AI Job Market & Salary Trends 2025 אשר מכיל אלפי רשומות המתארות מגוון רחב של משרות

בתחום ה-AI ממדינות רבות בעולם. הדאטה כולל פרטים כגון: תפקיד, שכר, רמת ניסיון, סוג העסקה, גודל

החברה, רמת עבודה מרחוק, מיקום גיאוגרפי ועוד.

נרצה לענות על מספר שאלות משנה נוספות שיובילו אותנו בפרויקט:

1. האם קיימים הבדלים גיאוגרפיים מובהקים באימוץ AI ברחבי העולם?
2. ניתוח השוואתי רב-ממדי של תפקידי עבודה בתחום ה-AI
3. השוואה בין רמת הביקוש למיומנויות שונות בשוק העבודה בתחום ה-AI לבין רמת השכר המוצעת עבורן.
4. האם קיימים הבדלים משמעותיים בשכר בתחום ה-AI בהתאם לתפקיד, אזור גיאוגרפי ורמת ניסיון?

חלק שני : ניתוח נתונים ומטלות

הדאטה שבחרנו לעבוד איתה מכילה 30,000 רשומות ו-19 פיצ'רים, כאשר כל רשומה מייצגת מודעת עבודה אחת

בתחום ה-AI. לכל פריט יש סט של מאפיינים המתארים את התפקיד, השכר, דרישות הניסיון, המיקום והתנאים

הנלווים.

מיפוי הנתונים בהתאם לטיפולוגיה האבסטרקטית של Munzner:

Dataset Type: הנתונים הינם בטבלאות עם מפתח אחד לכל item ולכן הסוג הוא FLAT TABLE

Data types:

- כל item בנתונים שלנו מייצג מודעת דרושים של משרה בתחום ה-AI.

- Attributes בטבלה מגוון שדות קטגוריאליים, אורדינליים, כמותיים.

Attribute	Type	Meaning	Cardinality
job_id (key)	Ordinal	מספר הרשומה	30,000
job_title	Categorical	שם המשרה	20
salary_usd	Quantitative	המשכורות בדולרים	27,548
salary_currency	Categorical	המטבע של המשכורת המקומית	8

experience_level	Ordinal	רמת ותק	4
employment_type	Ordinal	היקף המשרה	4
company_location	Categorical	המדינה בה המשרה מוצעת	20
company_size	Ordinal	גודל החברה	3
employee_residence	Categorical	מקום מגורי העובד	50
remote_ratio	Quantitative	רמת התאמת המשרד לעבודה מרחוק	3
required_skills	Categorical	כמעט כל רשומה עם רשימת מיומנויות ייחודית	25,702
education_required	Ordinal	High School / Bachelor / Master / PhD	4
years_experience	Quantitative	טווח שנות ניסיון שונים (20–0)	20
industry	Categorical	ענפים שונים	15
posting_date	Ordinal	תאריכי פרסום שונים	486
application_deadline	Quantitative	דדליין שונים	543
job_description_length	Quantitative	ערכים שונים (אורך טקסט המשרה)	2,000
benefits_score	Quantitative	ערכי ציון הטבות שונים	51
company_name	Categorical	חברות שונות	16

מטלה 1: האם קיימים הבדלים גיאוגרפיים מובהקים באימוץ AI ברחבי העולם?

במונחים של התחום: המשימה בוחנת כיצד אימוץ והשפעת טכנולוגיות AI משתנים בין מדינות ואזורים שונים בעולם. הבדלים רגולטוריים, תרבותיים וכלכליים עשויים להשפיע על קצב האימוץ, סוגי השימוש וההשפעה על עובדים ופרודוקטיביות.

המשימה משלבת בין נתוני הדאטה לבין מידע גאוגרפי, ומטרתה לזהות דפוסים מרחביים באימוץ AI.

במונחים של munzner:

Action	Sub action	Target	Explanation
analyze → consume	present	distribution	הצגת התפלגות גלובלית של משרות ועובדים על גבי מפה גיאוגרפית.
search	locate	outliers	זיהוי מדינות בעלות פעילות חריגה (גבוהה מאוד או נמוכה) ביחס לשאר העולם.
query	compare	trends	השוואה בין מדינות שונות (Comparator Mode) לבחינת הבדלים בשכר ובסוגי התפקידים.
query	summarize	patterns	קבלת תמונת מצב כללית על ריכוזי התעשייה (למשל: צפון אמריקה מול אירופה).

מטלה 2: ניתוח השוואתי רב-ממדי של תפקידי עבודה בתחום ה-AI

במונחים של התחום: מטלה זו עוסקת בניתוח השוואתי רב-ממדי של תפקידי עבודה בתחום ה-AI, במטרה לאמוד את ה"אטרקטיביות" של משרות שונות מנקודות מבט מגוונות ולא על סמך מדד בודד. באמצעות ויזואליזציה אינטראקטיבית, המערכת מאפשרת למשתמשים לחקור קשרים ודפוסים בין תפקידים (כגון Data Scientist או ML Engineer) על בסיס חמישה מאפייני מפתח: שכר, ניסיון נדרש, היקף ביקוש, הטבות ופיזור תעשייתי. הניתוח נועד לחשוף פערים וחריגים בשוק, למשל, תפקידים בעלי שכר גבוה אך דרישות סף גבוהות או היצע משרות נמוך ובכך לספק למשתמש תמונה הוליסטית ומורכבת המסייעת בקבלת החלטות קריירה מושכלות.

במונחים של munzner:

Action	Sub-action	Target	Explanation
analyze → consume	present	distribution	הצגת הפיזור של תפקידי AI שונים על פני מספר מדדים מרכזיים (שכר, ביקוש, ניסיון נדרש, הטבות ותעשייה),

			במטרה להבין את המיקום היחסי של כל תפקיד בשוק העבודה.
search	locate	outliers	זיהוי תפקידי AI חריגים, כגון תפקידים עם שכר גבוה במיוחד אך ביקוש נמוך, או תפקידים עם ביקוש רחב אך תגמול ושורת הטבות מצומצמים.
query	compare	trends	השוואת תפקידי AI שונים תחת מדדים שונים, למשל השוואה בין תפקידי מחקר לתפקידי פיתוח מבחינת שכר וניסיון נדרש, או בין תפקידי מוצר לתפקידי תשתית מבחינת ביקוש ופיזור תעשייתי.
query	summarize	patterns	זיהוי דפוסים כלליים בשוק העבודה, כגון תפקידים המאופיינים באטרקטיביות גבוהה במספר מדדים במקביל לעומת תפקידים המצטיינים בממד אחד בלבד אך חלשים באחרים.

מטלה 3 : השוואה בין רמת הביקוש למיומנויות שונות בשוק העבודה בתחום ה-AI לבין רמת השכר המוצעת עבורו.

במונחים של התחום: המשימה בוחנת את הקשר בין ביקוש למיומנויות טכנולוגיות שונות בתחום ה-AI (כפי שמשתקף בכמות המשרות הדורשות כל מיומנות) לבין רמת השכר השנתית הממוצעת. ניתוח זה מאפשר לזהות מיומנויות מבוקשות במיוחד, מיומנויות מתגמלות כלכלית, והאם קיים פער בין ביקוש גבוה לתגמול כספי. מטרת המשימה לחשוף דפוסים ומיומנויות חריגות בשוק העבודה.

במונחים של munzner:

Action	Sub action	Target	Explanation
analyze → consume	present	distribution	הצגת פיזור המיומנויות על פני צירי הביקוש (מספר משרות) והשכר השנתי, במטרה להבין את המיקום היחסי של כל מיומנות בשוק העבודה.
search	locate	outliers	זיהוי מיומנויות חריגות, כגון מיומנויות עם שכר גבוה אך ביקוש נמוך או מיומנויות עם ביקוש גבוה אך שכר ממוצע נמוך יחסית

query	compare	trends	השוואת מגמות בין קבוצות מיומנויות שונות, כגון שפות תכנות, כלי תשתית וטכנולוגיות למידת מכונה, כדי לבחון כיצד הביקוש משפיע על רמת השכר.
query	summarize	patterns	זיהוי דפוסים כלליים, למשל מיומנויות הממוקמות מעל החציון הן בביקוש והן בשכר, לעומת מיומנויות הנמצאות בפער בין ביקוש גבוה לתגמול כספי נמוך.

מטלה 4: האם קיימים הבדלים משמעותיים בשכר בתחום ה-AI בהתאם לתפקיד, אזור גיאוגרפי ורמת ניסיון?

במונחים של התחום: במטלה זו נבחן כיצד מאפיינים מקצועיים ותעסוקתיים משפיעים על רמות השכר בשוק העבודה בתחום ה-AI.

באמצעות ניתוח נתוני שכר, נבחן את הקשר בין שכר לבין תפקיד, אזור גיאוגרפי ורמת ניסיון, ונבדוק האם קיימים פערים מובהקים בין קבוצות שונות.

במונחים של munzner:

Action	Sub action	Target	Explanation
analyze → consume	present	correlation	הצגת הקשר בין שכר לבין תפקיד, אזור גיאוגרפי ורמת ניסיון באמצעות ויזואליזציות השוואתיות
search	locate	outliers	זיהוי משרות או אזורים עם שכר חריג במיוחד (גבוה או נמוך)
query	compare	trends	השוואת מגמות שכר בין תפקידים שונים, אזורים שונים ורמות ניסיון שונות
query	summarize	patterns	זיהוי דפוסים כלליים בשוק העבודה של AI, כגון שילובי תפקיד אזור ניסיון המניבים שכר גבוה במיוחד

חלק שלישי: עיצובים חלופיים

נציג את העיצובים החלופיים עבור כל מטלה שהצגנו:

מטלה 1: האם קיימים הבדלים גיאוגרפיים מובהקים באימוץ AI ברחבי העולם?

במונחים של התחום: המשימה בוחנת כיצד אימוץ והשפעת טכנולוגיות AI משתנים בין מדינות ואזורים שונים בעולם. הבדלים רגולטוריים, תרבותיים וכלכליים עשויים להשפיע על קצב האימוץ, סוגי השימוש וההשפעה על עובדים ופרודוקטיביות.

המשימה משלבת בין נתוני הדאטה לבין מידע גאוגרפי, ומטרתה לזהות דפוסים מרחביים באימוץ AI.

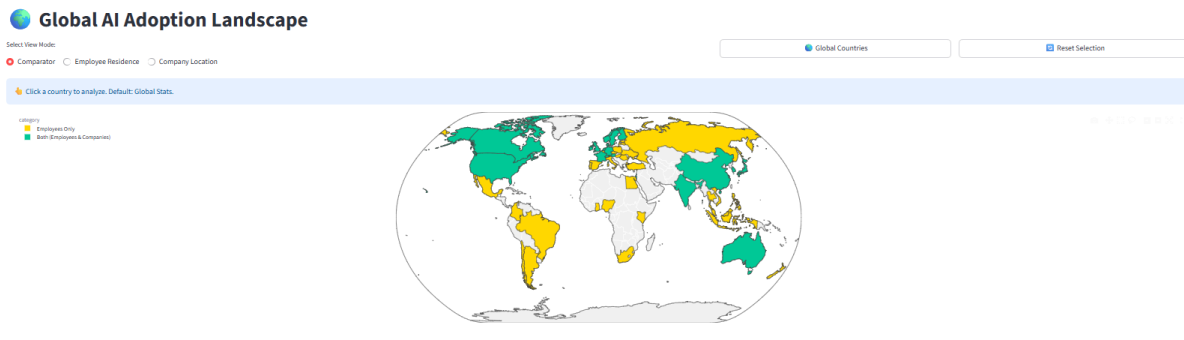
במונחים של munzner:

Action	Sub action	Target	Explanation
analyze → consume	present	distribution	הצגת התפלגות גלובלית של משרות ועובדים על גבי מפה גיאוגרפית.
search	locate	outliers	זיהוי מדינות בעלות פעילות חריגה (גבוהה מאוד או נמוכה) ביחס לשאר העולם.
query	compare	trends	השוואה בין מדינות שונות (Comparator Mode) לבחינת הבדלים בשכר ובסוגי התפקידים.
query	summarize	patterns	קבלת תמונת מצב כללית על ריכוזי התעשייה (למשל: צפון אמריקה מול אירופה).

ב. סקיצה:

חלופה 1:

- תיאור החלופה: הדשבורד מורכב ממפת עולם אינטראקטיבית (Choropleth Map).
- צבע המדינה משתנה בהתאם לכמות המשרות/העובדים (סקאלת צבעים רציפה).
 - לחיצה על מדינה במפה מסננת את הנתונים ומציגה בצד הדשבורד גרפים מפורטים עבור המדינה הנבחרת (Top Roles, Pay Ladder, Scatter Plot).
 - קיים מצב "Comparator" המאפשר בחירת שתי מדינות והשוואה ישירה ביניהן בגרפים מקבילים.



ג. תיאור החלופה במונחים של Marks & Channels

● Marks:

- Areas (שטחים): גבולות גיאוגרפיים של מדינות (2D Area).

● Channels:

- Color Saturation (רוויית צבע): משתנה כמותי (מספר משרות/עובדים). כהה יותר = יותר משרות.

- Spatial Region (מיקום מרחבי): משתנה קטגוריאלי (זהות המדינה).

- Interaction (אינטראקציה): Select/Click לביצוע סינון (Filter).

ד. אקספרסיביות ואפקטיביות

● אקספרסיביות: החלופה מציגה את כל המדינות הקיימות בדאטה. השימוש במפה הוא ייצוג טבעי למידע גיאוגרפי. מצב ההשוואה (Comparator) חושף מידע השוואתי שלא ניתן לראות במבט יחיד על מפה.

● אפקטיביות:

- ערוץ המיקום המרחבי הוא האפקטיבי ביותר לזיהוי מדינות.
- ערוץ הצבע (Color Saturation) פחות מדויק להשוואה כמותית עדינה (קשה להבחין בהבדל בין גוון 60% ל-70%), אך מצוין לזיהוי מגמות מקרו (איפה יש "הרבה" ואיפה "מעט").

ה. יתרונות וחסרונות

● יתרונות:

1. אינטואיטיביות: מפה היא הדרך המהירה ביותר למשתמש להבין הקשר גיאוגרפי ללא צורך בלימוד המערכת.
2. Overview first, details on demand: המפה מספקת תמונת על, והאינטראקציה מאפשרת צלילה לפרטים (Drill-down) מבלי להעמיס על המסך הראשוני.
3. הקשר מרחבי: מאפשרת לזהות דפוסים אזוריים (למשל, גוש מדינות באירופה עם אימוץ גבוה) שלא היו נראים בגרף עמודות רגיל.

● חסרונות:

1. Area Bias (הטיית שטח): מדינות גדולות פיזית (כמו רוסיה או קנדה) תופסות הרבה תשומת לב ויזואלית (דיו) גם אם הנתונים שלהן נמוכים, בעוד מדינות קטנות עם נתונים גבוהים (כמו ישראל או סינגפור) עלולות להיעלם.
2. דיוק נמוך בהשוואה: קשה להשוות ערכים כמותיים מדויקים באמצעות גוני צבע בלבד (האם צרפת כהה יותר מגרמניה?).

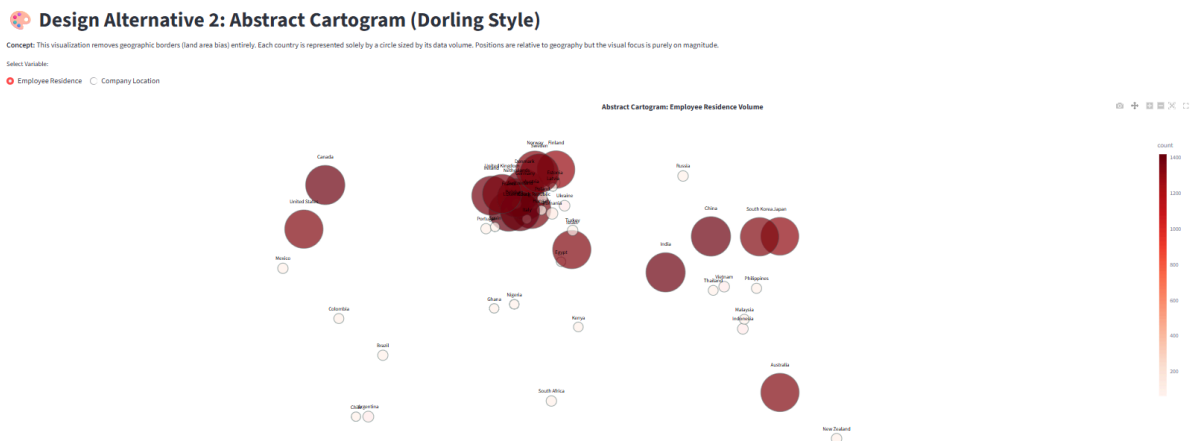
חלופה 2: קרטוגרמה מופשטת (Dorling Cartogram) - העיצוב שלא נבחר

חלופה זו מבוססת על הקוד השני ("Design Alternative 2"), המציג עיגולים המייצגים מדינות ללא גבולות גיאוגרפיים, כאשר גודל העיגול נקבע לפי כמות הנתונים.

א. תיאור המטלה (Task Abstraction)

Action	Sub action	Target	Explanation
analyze → consume	present	distribution	הצגת התפלגות גלובלית של משרות ועובדים על גבי מפה גיאוגרפית.
search	locate	outliers	זיהוי מדינות בעלות פעילות חריגה (גבוהה מאוד או נמוכה) ביחס לשאר העולם.
query	compare	trends	השוואה בין מדינות שונות (Comparator Mode) לבחינת הבדלים בשכר ובסוגי התפקידים.
query	summarize	patterns	קבלת תמונת מצב כללית על ריכוזי התעשייה (למשל: צפון אמריקה מול אירופה).

ב. סקיצה



ג. ניתוח Marks & Channels

1. Mark (סימן): נקודה/שטח (Point/Area). העיגול הוא היחידה הגרפית המייצגת מדינה.
2. Channel 1 (ערוץ גודל): שטח דו-ממדי (2D Area). שטח העיגול מקודד את המשתנה הכמותי (מספר המשרות). זהו הערוץ המרכזי והדומיננטי ביותר בחלופה זו.
3. Channel 2 (ערוץ מיקום): מיקום מרחבי (Spatial Position). המיקום הוא יחסי (טופולוגי) ולא אבסולוטי. הוא נועד לשמר את ה"מודל המנטלי" של מפת העולם (אירופה במרכז, ארה"ב משמאל) אך אינו מדויק גיאוגרפית.
4. Channel 3 (צבע - אופציונלי): ניתן להשתמש בצבע לקידוד מידע נוסף (כגון שכר ממוצע) או כ-Redundant Coding לחיזוק הקריאות.

ד. עקרונות האקספרסיביות והאפקטיביות

- אקספרסיביות: החלופה אקספרסיבית מאוד לנתוני Magnitude. היא מבטלת את "פקטור השקר" (Lie Factor) של מפות כורופלת (Choropleth), שבהן שטח גיאוגרפי ריק תופס נפח ויזואלי ומעוות את תפיסת המידע. בחלופה זו, רק מידע רלוונטי (Data) תופס מקום במסך (Ink).
- אפקטיביות:
 - יתרון: יעילה מאוד לזיהוי "מי השחקנים המרכזיים" ללא רעשי רקע.
 - חיסרון: על פי Stevens' Psychophysical Power Law, בני אדם מתקשים להעריך יחסי שטחים בצורה ליניארית (אנו נוטים להעריך שטח בחסר לעומת אורך). לכן, המשתמש יתקשה לדעת אם עיגול א' גדול בדיוק פי 2 מעיגול ב'. כמו כן, האפקטיביות של מטלת ה-Locate (איתור מדינה ספציפית) נפגעת משמעותית בהעדר גבולות מוכרים.

1. צדק ויזואלי (Fairness): מדינות עם השפעה גדולה בתחום ה-AI מקבלות ייצוג ויזואלי בולט, ללא קשר לשטחן הפיזי. ישראל, סינגפור ובריטניה יקבלו ייצוג הולם, בעוד שטחים עצומים וריקים (כמו סיביר או צפון קנדה) יעלמו ולא יסיחו את הדעת.
2. Data-Ink Ratio גבוה: הסרת הגבולות, הימים והרקע הגיאוגרפי ממקדת את המשתמש אך ורק בנתונים עצמם (De-cluttering).

חסרונות:

1. אובדן קונטקסט (Loss of Context): המשתמש מאבד את ההקשר הגיאוגרפי המדויק. קשה לזהות מדינות שאינן בולטות במיוחד (עיגולים קטנים) ללא קווי מתאר של יבשת שיעזרו בהתמצאות.
2. עומס קוגניטיבי בזיהוי: ללא תוויות ברורות (Labels) על כל עיגול, המשתמש יתקשה להבין איזו מדינה מיוצגת על ידי איזה עיגול, למעט המדינות המוכרות ביותר במיקומן המשוער.

הצדקת הבחירה בחלופה הסופית:

- אינטואיטיביות ומוכרות: החלופה מציגה את הנתונים על גבי מודל מנטלי מוכר (מפת העולם), מה שמאפשר למשתמש להתמצא במרחב באופן מיידי ללא צורך בלמידת ייצוג חדש ומופשט.
 - איתור מהיר (Locate): ניתן לאתר מדינות ספציפיות בקלות על בסיס צורתן ומיקומן הגיאוגרפי, בניגוד לקרטוגרמה המופשטת שבה המיקום היחסי משתבש והזיהוי הופך למאמץ קוגניטיבי.
 - שימור הקשר מרחבי: החלופה משמרת את הטופולוגיה (שכנות בין מדינות), מה שמאפשר זיהוי דפוסים ומגמות אזוריות (למשל: ריכוז גבוה באירופה) בצורה טובה יותר מאשר בועות מנותקות.
 - גישה לנתונים מדויקים: אינטראקציית hover מאפשרת גישה לערכים המספריים המדויקים, ובכך מפצה על החיסרון של ערוץ הצבע (Color Saturation) בהשוואה כמותית עדינה.
 - עקרונות Munzner: החלופה מממשת את עקרון השימוש ב-Spatial Region (מיקום מרחבי) כערוץ האפקטיבי ביותר לנתונים קטגוריאליים (זהות המדינה), תוך העדפת קלות הזיהוי על פני דיוק בהשוואת שטחים (שבו בני אדם מתקשים לפי חוק Stevens).
 - מניעת עומס ויזואלי: המפה מספקת "עוגן" ויזואלי ברור, בעוד שאוסף רב של עיגולים מרחפים (בחלופה השנייה) עלול ליצור עומס ויזואלי ובלבול כאשר מדובר במספר רב של מדינות.
- הקרטוגרמה המופשטת (Dorling Cartogram) נשארת חלופה תקפה להצגת סדרי גודל נקיים ללא "הטיית שטח" (Area Bias), אך לא נבחרה כמימוש הסופי בשל הפגיעה המשמעותית ביכולת ההתמצאות ובזיהוי המהיר של מדינות ללא תוויות טקסט מרובות.

מטלה 2: ניתוח השוואתי רב-ממדי של תפקידי עבודה בתחום ה-AI

מטלה זו עוסקת בניתוח השוואתי רב-ממדי של תפקידי עבודה בתחום ה-AI, במטרה לאמוד את ה"אטרקטיביות" של משרות שונות מנקודות מבט מגוונות ולא על סמך מדד בודד. באמצעות ויזואליזציה אינטראקטיבית, המערכת מאפשרת למשתמשים לחקור קשרים ודפוסים בין תפקידים (כגון Data Scientist או ML Engineer) על בסיס חמישה מאפייני מפתח: שכר, ניסיון נדרש, היקף ביקוש, הטבות ופיזור תעשייתי. הניתוח נועד לחשוף פערים וחריגים בשוק, למשל, תפקידים בעלי שכר גבוה אך דרישות סף גבוהות או היצע משרות נמוך ובכך לספק למשתמש תמונה הוליסטית ומורכבת המסייעת בקבלת החלטות קריירה מושכלות.

Action	Sub-action	Target	Explanation
analyze → consume	present	distribution	הצגת הפיזור של תפקידי AI שונים על פני מספר מדדים מרכזיים (שכר, ביקוש, ניסיון נדרש, הטבות ותעשייה), במטרה להבין את המיקום היחסי של כל תפקיד בשוק העבודה.
search	locate	outliers	זיהוי תפקידי AI חריגים, כגון תפקידים עם שכר גבוה במיוחד אך ביקוש נמוך, או תפקידים עם ביקוש רחב אך תגמול ושורת הטבות מצומצמים.
query	compare	trends	השוואת תפקידי AI שונים תחת מדדים שונים, למשל השוואה בין תפקידי מחקר לתפקידי פיתוח מבחינת שכר וניסיון נדרש, או בין תפקידי מוצר לתפקידי תשתית מבחינת ביקוש ופיזור תעשייתי.
query	summarize	patterns	זיהוי דפוסים כלליים בשוק העבודה, כגון תפקידים המאופיינים באטרקטיביות גבוהה במספר מדדים במקביל לעומת תפקידים המצטיינים במדד אחד בלבד אך חלשים באחרים.

חלופת עיצוב 1: תרשים עמודות מקובץ (Grouped Bar Chart) עם מפת חום משלימה

א. המטלה עליה עונה החלופה

במונחים אבסטרקטיים:

השוואה, זיהוי חריגים וסיכום דפוסים (Compare, Locate, Summarize)
במונחי התחום:

השוואת תפקידי AI על פני מספר מדדים מרכזיים לצורך הערכת האטרקטיביות היחסית שלהם בשוק העבודה.

ב. סקיצה

הסקיצה מציגה את תפקידי העבודה על הציר האופקי, ולצדם קבוצות של עמודות. כל עמודה מייצגת מדד שונה (כגון שכר, היקף משרות או ניסיון נדרש), כאשר הערכים מנורמלים לטווח אחיד לצורך השוואה בין מדדים בעלי יחידות שונות. מתחת לתרשים מוצגת טבלה משלימה עם קידוד צבעים, המציגה את הערכים הגולמיים של אותם מדדים.

Multidimensional Comparison of AI Job Roles

Compare AI job roles across disparate metrics (Salary, Frequency, Remote availability) using Min-Max Normalization.

Select Metrics for Comparison

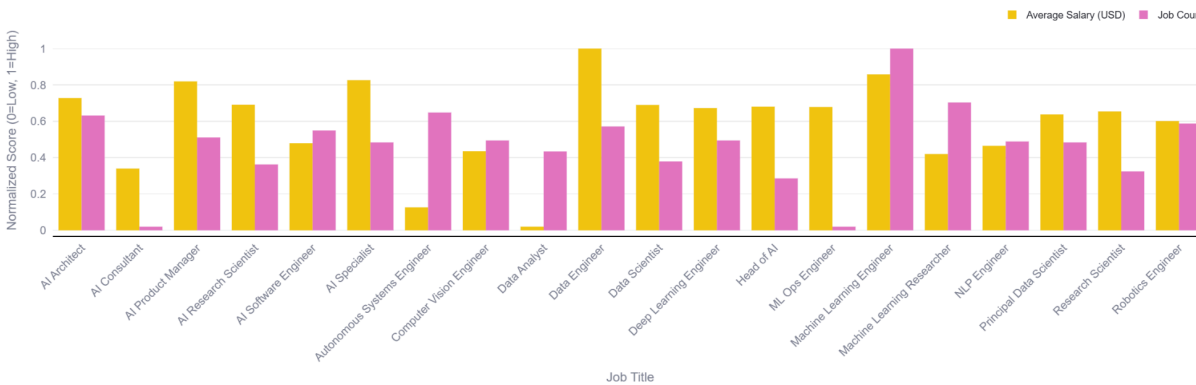
☒ Average Salary (USD) ☒ Job Count ☐ Years of Experience ☐ Benefits Score ☐ Remote Work Ratio (%)

Chart Sort Settings

Sort Chart By:

Name (A-Z)

Normalized Multidimensional Comparison (Sorted by: Name (A-Z))



Underlying Data (Raw Values)

Search by Job Title

Type job name...

job_title	Average Salary (USD)	Job Count
AI Architect	119,725.43	1,529.00
AI Consultant	116,733.32	1,417.00
AI Product Manager	120,434.36	1,507.00
AI Research Scientist	119,444.34	1,480.00
AI Software Engineer	117,807.11	1,514.00
AI Specialist	120,487.82	1,502.00
Autonomous Systems Engineer	115,083.59	1,532.00
Computer Vision Engineer	117,467.19	1,504.00
Data Analyst	114,109.41	1,493.00
Data Engineer	121,828.42	1,518.00
Data Scientist	119,434.51	1,483.00
Deep Learning Engineer	119,301.77	1,504.00
Head of AI	119,360.61	1,466.00
ML Ops Engineer	119,346.68	1,414.00

ג. Marks & Channels

1. תרשים העמודות:

Marks:

עמודות (מלבנים)

Channels:

- מיקום אופקי - תפקיד עבודה
- גובה העמודה - ערך כמותי מנורמל
- צבע (קטגוריאלי) -סוג המדד
- קיבוץ - שיוך מספר מדדים לאותו תפקיד

2. מפת החום (טבלת נתונים משלימה)

Marks:

תאי טבלה

Channels:

- מיקום בשורות - תפקידי עבודה
- מיקום בעמודות - מדדים
- צבע רציף - גודל הערך הגולמי של המדד
- טקסט- הערך המספרי המדויק

ד. אקספרסיביות ואפקטיביות

אקספרסיביות:

כל ערוץ ויזואלי מייצג משתנה נתונים רלוונטי באופן ישיר וברור, ללא הצגת מידע שאינו קיים בנתונים. בתרשים העמודות, הנרמול מאפשר ייצוג הוגן של מדדים בעלי יחידות שונות, בעוד שמפת החום מציגה את הערכים הגולמיים באופן ישיר.

אפקטיביות:

השימוש במיקום ובגובה העמודות מאפשר השוואה מדויקת בין תפקידים. הנרמול תומך בהשוואה רב-ממדית, כאשר אינטראקציית ה-hovert מספקת גישה לערכים האבסולוטיים ומפצה על אובדן קנה המידה המקורי. מפת החום משלימה את התרשים בכך שהיא מאפשרת זיהוי מהיר של ערכים גבוהים ונמוכים ושומרת על דיוק מספרי.

ה. יתרונות וחסרונות

יתרונות:

- השוואה ברורה בין תפקידי AI על פני מספר מדדים
- זיהוי קל של דפוסים
- שילוב בין מבט כולל (תרשים מנורמל) לבין נתונים גולמיים (מפת חום)

- אפקטיביות גבוהה בהשוואות כמותיות

חסרונות:

- הנירמול עלול להסתיר הבדלים אבסולוטיים ללא שימוש באינטראקציה
- עומס ויזואלי עולה עם ריבוי מדדים
- מפת החום אפשרית לפי עמודה ולא לפי כל הטבלה(סקלה שונה לכל פיצר והערכים ללא נרמול)

חלופת עיצוב 2: תרשים קואורדינטות מקבילות (Parallel Coordinates)

א. המטלה עליה עונה החלופה

- אבסטרקטית:

השוואה רב־ממדית וזיהוי דפוסים וחריגים
(Compare, Summarize, Locate outliers)

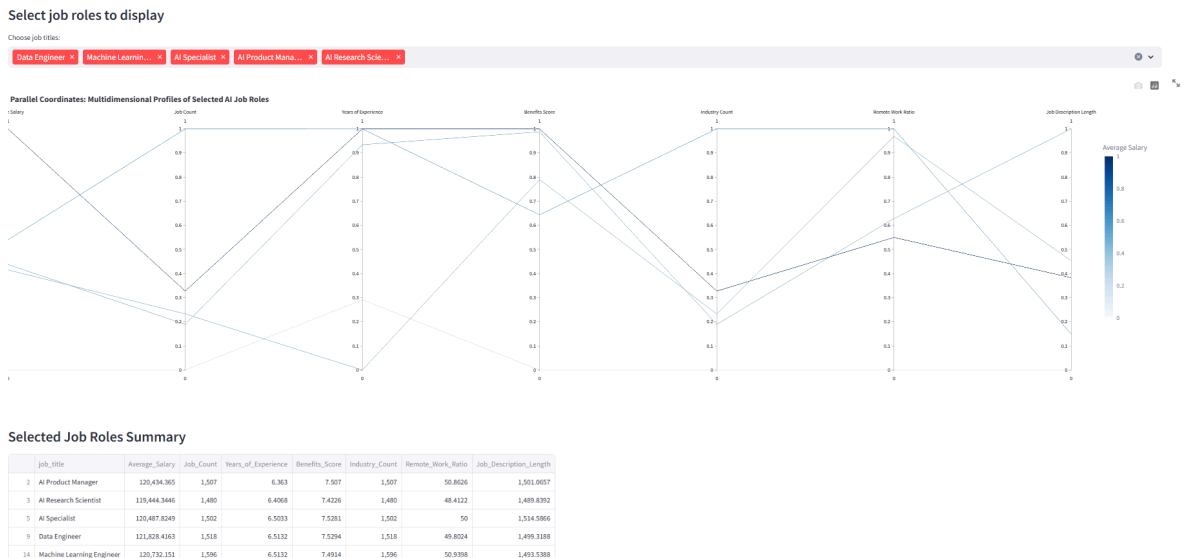
- במונחי התחום:

השוואת תפקידי AI על פני שכר, ניסיון, ביקוש, הטבות, פיזור תעשייתי ועבודה מרחוק.

ב. סקיצה

הסקיצה כוללת מספר צירים אנכיים מקבילים, כאשר כל ציר מייצג מדד כמותי אחד.

כל תפקיד עבודה מיוצג על ידי קו המחבר בין ערכיו על פני הצירים השונים.



ג. Marks & Channels

Marks:

- קווים (קו אחד לכל תפקיד)
- צירים אנכיים

Channels:

- מיקום לאורך הציר - ערך כמותי
- צבע רציף - שכר ממוצע

- סדר הצירים - סוג המדד

ד. אקספרסיביות ואפקטיביות

- אקספרסיביות:

כל ערוץ ויזואלי מייצג מאפיין נתונים רלוונטי, ללא קידוד מיותר.

- אפקטיביות:

מיקום הוא ערוץ אפקטיבי להשוואת ערכים, אך חפיפות בין קווים מקשות על השוואה מדויקת בין תפקידים בודדים.

בנוסף, זיהוי של תפקיד מסוים דורש אינטראקציה משלימה או טבלה נלווית.

ה. יתרונות וחסרונות

יתרונות

- מאפשר הצגה של מספר רב של ממדים בו־זמנית

- מתאים לזיהוי דפוסים כלליים ויחסי trade-off בין מדדים

חסרונות

- קושי בזיהוי חד־משמעי של תפקיד ספציפי

- חפיפות רבות יוצרות עומס ויזואלי

- דיוק נמוך בהשוואה נקודתית של ערכים

לאחר בחינת שתי החלופות, נבחרה חלופת תרשים העמודות כמימוש הסופי של הפרויקט מכיוון:

- מציגה נתונים באופן פשוט ונוח למשתמש

- ניתן להבין על איזה משרה מדובר בקלות

- אינטראקציית hover מאפשרת גישה לערכים מדויקים ומפחיתה את החיסרון של הנרמול

- החלופה מספקת תמיכה טובה יותר במשימות השוואה בין תפקידים בודדים

- היא עומדת טוב יותר בעקרונות האפקטיביות של Munzner באמצעות שימוש במיקום ובאורך

- בהשוואה לקואורדינטות מקבילות, מתקבל ייצוג פחות עמוס וברור יותר למשתמש

תרשים הקואורדינטות המקבילות נשאר חלופה תקפה לחקירה ראשונית, אך לא נבחר כמימוש הסופי בשל מגבלותיו בזיהוי והשוואה מדויקת של תפקידים.

מטלה 3 : השוואה בין רמת הביקוש למיומנויות שונות בשוק העבודה בתחום ה-AI לבין רמת השכר המוצעת עבורן.

א. תיאור המטלה במונחים אבסטרקטיים ובמונחי התחום עבור שתי החלופות:

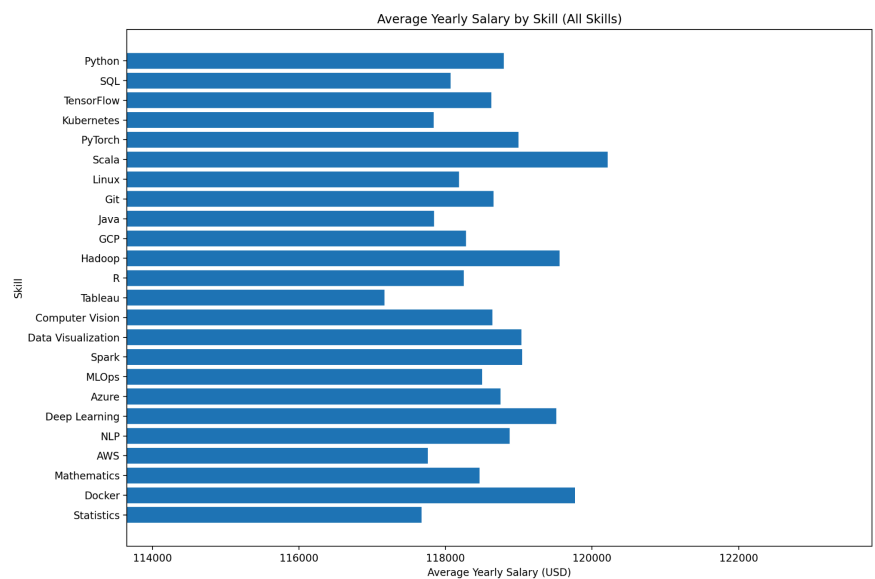
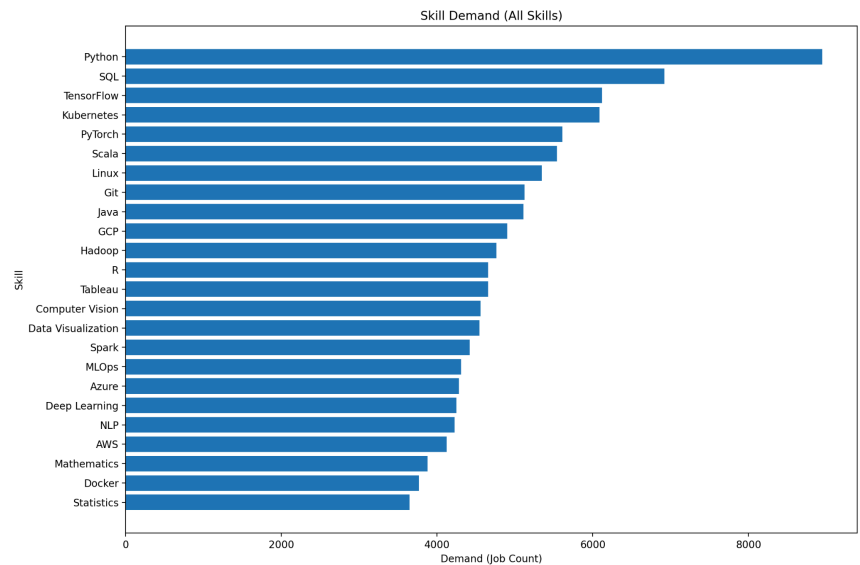
במונחים של התחום: המשימה בוחנת את הקשר בין ביקוש למיומנויות טכנולוגיות שונות בתחום ה-AI (כפי שמשתקף בכמות המשרות הדורשות כל מיומנות) לבין רמת השכר השנתית הממוצעת. ניתוח זה מאפשר לזהות מיומנויות מבוקשות במיוחד, מיומנויות מתגמלות כלכלית, והאם קיים פער בין ביקוש גבוה לתגמול כספי. מטרת המשימה לחשוף דפוסים ומיומנויות חריגות בשוק העבודה.

במונחים של munzner:

Action	Sub action	Target	Explanation
analyze → consume	present	distribution	הצגת פיזור המיומנויות על פני צירי הביקוש (מספר משרות) והשכר השנתי, במטרה להבין את המיקום היחסי של כל מיומנות בשוק העבודה.
search	locate	outliers	זיהוי מיומנויות חריגות, כגון מיומנויות עם שכר גבוה אך ביקוש נמוך או מיומנויות עם ביקוש גבוה אך שכר ממוצע נמוך יחסית
query	compare	trends	השוואת מגמות בין קבוצות מיומנויות שונות, כגון שפות תכנות, כלי תשתית וטכנולוגיות למידת מכונה, כדי לבחון כיצד הביקוש משפיע על רמת השכר.
query	summarize	patterns	זיהוי דפוסים כלליים, למשל מיומנויות הממוקמות מעל החציון הן בביקוש והן בשכר, לעומת מיומנויות הנמצאות בפער בין ביקוש גבוה לתגמול כספי נמוך.

ניתוח החלופה הראשונה:

ב. סקיצה של החלופה:



תיאור כללי של החלופה:

חלופה זו מציגה את הנתונים באמצעות שני גרפי Bar אופקיים נפרדים:

- גרף ראשון: ציר X - ביקוש (מספר משרות), ציר Y - מיומנויות
- גרף שני: ציר X - שכר שנתי ממוצע, ציר Y - מיומנויות
- הגרפים מוצגים זה מתחת לזה ומשתמשים באותו סדר מיומנויות לצורך השוואה ויזואלית. הגרפים הנ"ל ללא אינטראקטיביות, בעלי ייצוג סטטי.

ג. תיאור החלופה במונחים של Marks & Channels:

Marks:

- Bars (מלבנים)

Channels:

- Position (אורך הבר על ציר X): משתנה כמותי (Demand / Salary)

- Position (ציר Y): משתנה קטגורי (Skill)

- Color (אחיד): אינו נושא מידע, משמש להפרדה בלבד

ד. אקספרסיביות ואפקטיביות

אקספרסיביות:

- החלופה מייצגת רק משתנים קיימים (מיומנות, ביקוש, שכר)

- אין הצגה של מידע שלא קיים בדאטה

אפקטיביות:

- שימוש ב-Position לאורך ציר משותף הוא ערוץ מדויק להשוואה כמותית.

- מאפשר השוואה טובה בתוך כל משתנה בנפרד (מי יותר מבוקש / מי מתוגמל יותר).

- החיסרון- ההשוואה בין ביקוש לשכר היא השוואה חוצת-גרפים, ולכן דורשת מאמץ קוגניטיבי (השוואה בזיכרון)

ה. יתרונות וחסרונות:

יתרונות:

- השימוש בערוץ position לאורך ציר משותף מאפשר השוואה מדויקת של גדלים כמותיים (magnitude),

בהתאם לדירוג האפקטיביות של Munzner, שבו position נחשב לערוץ המדויק ביותר להשוואה כמותית.

- ההפרדה לשני גרפים מאפשרת מיקוד בכל משתנה בנפרד (ביקוש או שכר), ומונעת ערבוב בין משתנים בעלי יחידות מידה שונות.

- הייצוג הקטגורי על ציר ה-Y תומך בהשוואה ישירה בין מיומנויות רבות, גם כאשר מספר הקטגוריות גבוה יחסית.

חסרונות:

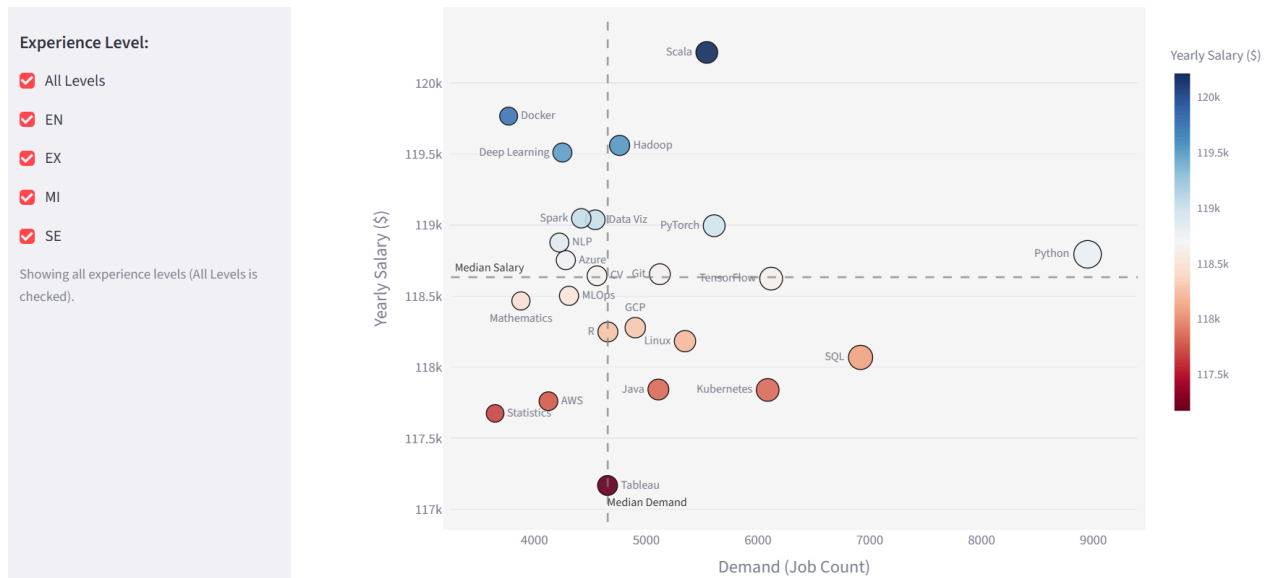
- מאחר שכל גרף מייצג משתנה אחד בלבד, אין קידוד משותף לשני משתנים, ולכן לא ניתן לזהות קשרים או יחסים בין ביקוש לשכר באמצעות ערוצים ויזואליים, אלא רק באמצעות השוואה קוגניטיבית בין גרפים נפרדים.

- היעדר שימוש בערוצים נוספים כגון color או size מגביל את יכולת הוויזואליזציה להעביר מידע רב-ממדי.

- זיהוי חריגים (למשל מיומנות עם ביקוש נמוך אך שכר גבוה) אינו נתמך ישירות על ידי הערוצים הוויזואליים, ודורש קריאה והשוואה ידנית בין שני הגרפים.

ניתוח החלופה השניה:

ב. סקיצה של החלופה:



תיאור כללי של החלופה:

חלופה זו מציגה את כל המידע בגרף אחד:

- כל נקודה מייצגת מיומנות
- מיקום אופקי - רמת ביקוש
- מיקום אנכי - שכר שנתי ממוצע
- צבע הנקודה - שכר שנתי
- גודל הנקודה - רמת ביקוש
- קווי חציון מחלקים את המרחב לארבעה אזורים

הגרף כולל אינטראקטיביות: סינון לפי רמות ניסיון.

ג. תיאור החלופה במונחים של Marks & Channels:

Marks:

- Points (עיגולים)

Channels:

- Position X: ביקוש (Quantitative)
- Position Y: שכר שנתי ממוצע (Quantitative)
- Size: ביקוש (redundant encoding)
- Color (Sequential): שכר
- Text: תוויות למיומנויות נבחרות
- Interaction: בחירה, סינון - filtering, selection

ד. אקספרסיביות ואפקטיביות

אקספרסיביות:

- כל ערוץ מייצג משתנה משמעותי
- אין קידוד כפול שמטעה (size מחזק את demand)

אפקטיביות:

- Position על שני צירים תומך בהשוואה מדויקת בין משתנים
- שילוב color + position מאפשר זיהוי דפוסים מהיר
- קווי חציון יוצרים חלוקה סמנטית ברורה של המרחב
- אינטראקטיביות תומכת ב-overview → filter → details on demand

ה. יתרונות וחסרונות:

יתרונות:

- השימוש בערוץ position על שני צירים כמותיים מאפשר השוואה ישירה בין ביקוש לשכר באותו מרחב ויזואלי, ותומך בזיהוי קשרים, דפוסים ויחסים בין משתנים.
- קידוד הביקוש הן באמצעות מיקום בציר X והן באמצעות size מספק חיזוק ויזואלי (redundant encoding), המקל על זיהוי מיומנויות בעלות ביקוש גבוה במיוחד.
- קידוד השכר באמצעות מיקום בציר Y ובאמצעות צבע בסקאלה רציפה מאפשר הבחנה הדרגתית בין רמות שכר שונות, גם כאשר ההבדלים המספריים ביניהן קטנים יחסית.
- קווי החציון מוסיפים מבנה מרחבי לנתונים ומאפשרים חלוקה סמנטית של המרחב לארבעה אזורים, ללא צורך בהוספת משתנים חדשים.
- האינטראקטיביות (סינון לפי רמת ניסיון) מאפשרת הפחתת עומס ויזואלי והתמקדות בתתי-אוכלוסיות, בהתאם לעקרון overview → filter → details on demand.

חסרונות:

- שימוש במספר ערוצים ויזואליים במקביל (position, size, color, text) מעלה את רמת העומס הוויזואלי ודורש ניהול זהיר של תוויות וחפיפות.
- ערוץ size נחשב לפחות מדויק מ-position להשוואה כמותית, ולכן אינו מתאים לביצוע השוואות מדויקות אלא בעיקר לזיהוי יחסי סדר וגודל.
- חפיפה בין נקודות עלולה להסתיר פריטים מסוימים, ודורשת פתרונות משלימים כגון תוויות סלקטיביות או אינטראקטיביות.

2. הצדקת הבחירה בחלופה הסופית:

החלופה שנבחרה למימוש בשלב האחרון של הפרויקט היא גרף ה-scatter האינטראקטיבי, המציג את הקשר בין רמת הביקוש למיומנויות שונות לבין רמת השכר השנתי הממוצע עבורן. בחירה זו נעשתה לאחר השוואה בין חלופה זו לבין חלופת גרפי ה-Bar, ובהתבסס על התאמת כל אחת מהן למטלות האנליטיות שהוגדרו בשלב הקודם.

חלופת גרפי ה-Bar מאפשרת השוואה מדויקת של כל משתנה בנפרד, באמצעות קידוד של ערכים כמותיים בערוץ position לאורך ציר משותף, אך אינה מאפשרת ייצוג משולב של ביקוש ושכר באותו מרחב ויזואלי. כתוצאה מכך, זיהוי קשרים, חריגים ודפוסים בין שני המשתנים מחייב השוואה קוגניטיבית בין גרפים נפרדים ואינו נתמך ישירות על ידי הערוצים הוויזואליים.

לעומת זאת, גרף ה-scatter שנבחר עושה שימוש ב-position על שני צירים כמותיים כדי לייצג בו־זמנית את הביקוש ואת השכר, ובכך מאפשר ניתוח ישיר של הקשר ביניהם. שילוב ערוצי color ו-size מחזק את ייצוג המידע ומאפשר זיהוי מהיר של דפוסים וחריגים, כגון מיומנויות בעלות ביקוש גבוה אך שכר נמוך, או מיומנויות בעלות שכר גבוה ביחס לביקוש. בנוסף, קווי החציון מחלקים את מרחב הנתונים לאזורים סמנטיים ומסייעים בסיווג המיומנויות לקבוצות משמעותיות.

יתרון נוסף של החלופה הנבחרת הוא האינטראקטיביות, המאפשרת סינון לפי רמות ניסיון שונות. יכולת זו תומכת בעקרון overview → filter → details on demand, ומאפשרת למשתמש לחקור את הנתונים באופן הדרגתי, תוך הפחתת עומס ויזואלי והתמקדות בתתי-אוכלוסיות רלוונטיות.

למרות ששימוש גרף ה-scatter מורכב יותר ודורש ניהול של עומס ויזואלי וחפיפות בין נקודות, יתרונותיו האנליטיים והיכולת שלו לתמוך במגוון רחב של מטלות משתמש מצדיקים את הבחירה בו כוויזואליזציה הסופית של משימה זו.

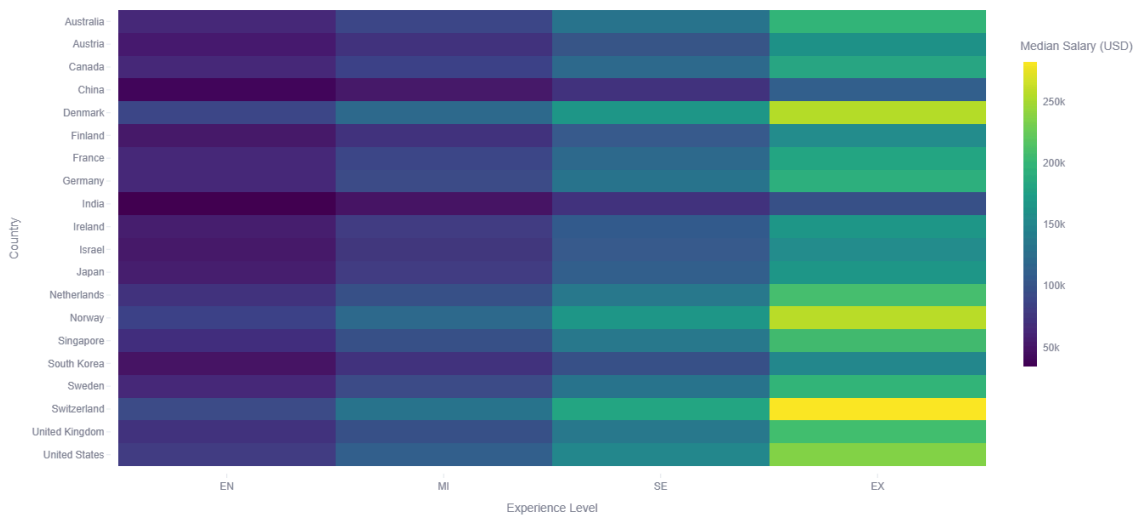
מטלה 4: האם קיימים הבדלים משמעותיים בשכר בתחום ה-AI בהתאם לתפקיד, אזור גיאוגרפי ורמת ניסיון?

א. תיאור המטלה במונחים אבסטרקטיים ובמונחי התחום עבור שתי החלופות:
במונחים של התחום: במטלה זו נבחן כיצד מאפיינים מקצועיים ותעסוקתיים משפיעים על רמות השכר בשוק העבודה בתחום ה-AI.
באמצעות ניתוח נתוני שכר ממאגר *Global AI Job Market and Salary Trends 2025*, נבחן את הקשר בין שכר לבין תפקיד, אזור גיאוגרפי ורמת ניסיון, ונבדוק האם קיימים פערים מובהקים בין קבוצות שונות.
במונחים של munzner:

Action	Sub action	Target	Explanation
analyze → consume	present	correlation	הצגת הקשר בין שכר לבין תפקיד, אזור גיאוגרפי ורמת ניסיון באמצעות ויזואליזציות השוואתיות
search	locate	outliers	זיהוי משרות או אזורים עם שכר חריג במיוחד (גבוה או נמוך)

query	compare	trends	השוואת מגמות שכר בין תפקידים שונים, אזורים שונים ורמות ניסיון שונות
query	summarize	patterns	זיהוי דפוסים כלליים בשוק העבודה של AI, כגון שילובי תפקיד אזור ניסיון המניבים שכר גבוה במיוחד

ניתוח החלופה הראשונה :
ב. סקיצה של החלופה:



ג. תיאור החלופה במונחים של Marks & Channels:
:Marks

- Area marks rectangular cells - כל תא במטריצה מייצג צירוף ייחודי של מדינה ו-רמת ניסיון, והשטח המלבני משמש כ-container לערך השכר החציוני. שימוש ב-area marks מתאים להצגת נתונים בטבלאות דו-ממדיות שבהן אין צורך להדגיש פיזור או ערכים בודדים, אלא קשרים בין קטגוריות.

:Channels

- Position (x) – Experience Level (categorical, ordered) - ציר ה-x מייצג את רמות הניסיון (EN → EX), סדר שמבוסס על היררכיה מושגית קיימת בנתונים. מיקום אופקי מאפשר קריאה השוואתית בין רמות ניסיון שונות עבור אותה מדינה.

- Position (y) – Country (categorical) | ציר ה-y מבדיל בין מדינות. מיקום אנכי מאפשר השוואה בין מדינות שונות בתוך אותה רמת ניסיון. שילוב שני צירי המיקום יוצר מטריצה שממקמת השוואות קטגוריאליות.
- Color luminance / saturation – Median Salary (quantitative, ordered) | הצבע מייצג ערך כמותי רציף (שכר חציוני), באמצעות colormap סדרתי (sequential). שימוש ב-luminance/saturation תומך בהבחנה בין ערכים נמוכים לגבוהים מבלי להוסיף עומס צורני.
- ד. אקספרסיביות ואפקטיביות
- אקספרסיביות:
- כל ערוץ מייצג מידע שקיים בדאטה בלבד, אין שימוש בצבע, מיקום או צורה כדי לרמוז על משתנה שלא קיים בטבלה המקורית.
- הצבע מייצג משתנה כמותי רציף באופן תקין, השכר החציוני מוצג באמצעות colormap רציף, ללא דיסקרטיזציה מלאכותית (כגון קבוצות "נמוך/גבוה"). בכך נמנעת יצירת גבולות מדומים שאינם קיימים בנתונים.
- הימנעות מערוצים נוספים שאינם נדרשים, אין שימוש בגודל, צורה, זווית או אנימציה - ערוצים שעלולים ליצור רמיזות שווא או להעמיס על הפרשנות.
- אין רמיזה להיררכיה, סיבתיות או דירוג' הוויזואליזציה מציגה קשרים והשוואות בלבד, מבלי להציע הסבר סיבתי או סדר עדיפות בין מדינות.

אפקטיביות:

- Color luminance מתאים להצגת מגמות ודפוסים כלליים, הצבע מאפשר לזהות במהירות אזורים "כהים" ו-"בהירים", ובכך לחשוף מגמות רוחביות והבדלים יחסיים.
- סידור מטריציוני תומך בהשוואות מרובות:
- קריאה אנכית: השוואת רמות ניסיון בתוך אותה מדינה
- קריאה אופקית: השוואת מדינות בתוך אותה רמת ניסיון/מבנה זה ממקסם את האפקטיביות של ערוץ המיקום.
- היעדר ערכים מספריים מפחית עומס קוגניטיבי, המשתמש מונחה לקריאה ברמת overview ולזיהוי דפוסים, ולא לביצוע lookup מדויק שאינו מטרת החלופה.
- מגבלה מודעת של הערוץ הצבעוני, צבע פחות אפקטיבי להשוואות מדויקות בין ערכים קרובים, אך מגבלה זו מוצדקת משום שהחלופה אינה מיועדת לדיוק מספרי אלא לניתוח מגמות.

ה. יתרונות וחסרונות:

יתרונות:

- זיהוי מהיר של מגמות רוחביות, לדוגמה: עלייה עקבית בשכר עם רמת ניסיון, או פערים ברורים בין מדינות.
- תמיכה חזקה במטלות ברמת overview החלופה מתאימה במיוחד ל- summarize, compare ו- identify patterns.

- ייצוג קומפקטי של דאטה רב קטגוריאלי, מאפשר הצגת מספר רב של מדינות ורמות ניסיון בגרף אחד מבלי ליצור עומס צורני.
- ניצול יעיל של ערוץ המיקום, המטריצה משתמשת במיקום אחד הערוצים האפקטיביים ביותר כדי לאפשר השוואות מרובות.

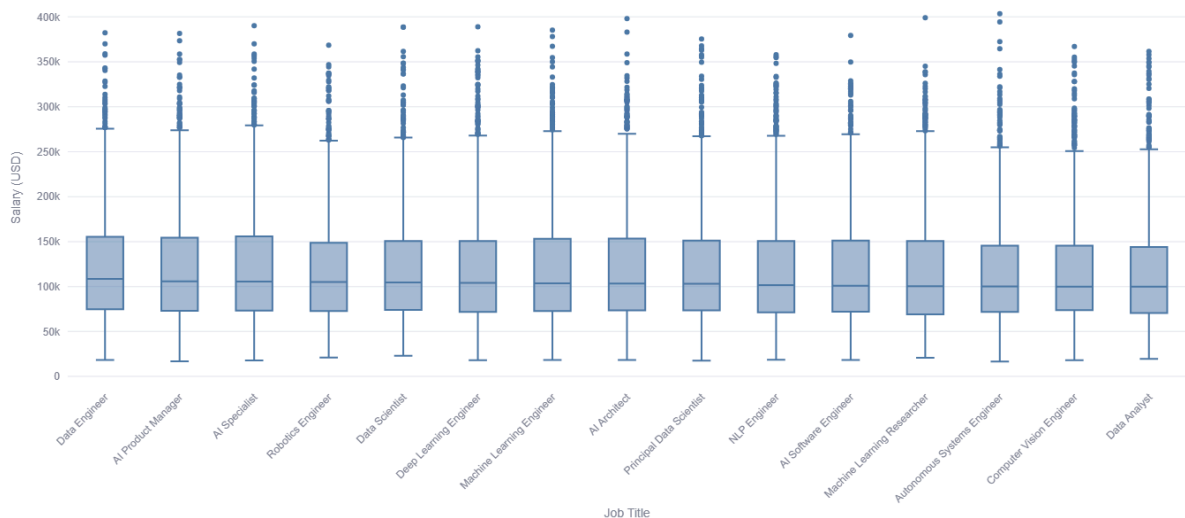
חסרונות:

- לא מתאים ל-lookup מדויק של ערכי שכר, היעדר מספרים מקשה על קריאה מדויקת של ערך ספציפי.
- הבחנה בין ערכים קרובים בצבע עשויה להיות מאתגרת במיוחד כאשר ההבדלים קטנים או כאשר הצופה רגיש פחות להבדלי luminance.
- תלות גבוהה בבחירת colormap, בחירה לא מוצלחת עלולה לפגוע בקריאות, בנגישות (עיוורון צבעים) ובהשוואתיות.
- אין הדגשה של חריגים נקודתיים, ערכים קיצוניים נטמעים בתוך המטריצה ואינם מובלטים כפי שקורה ב-scatter או boxplot.

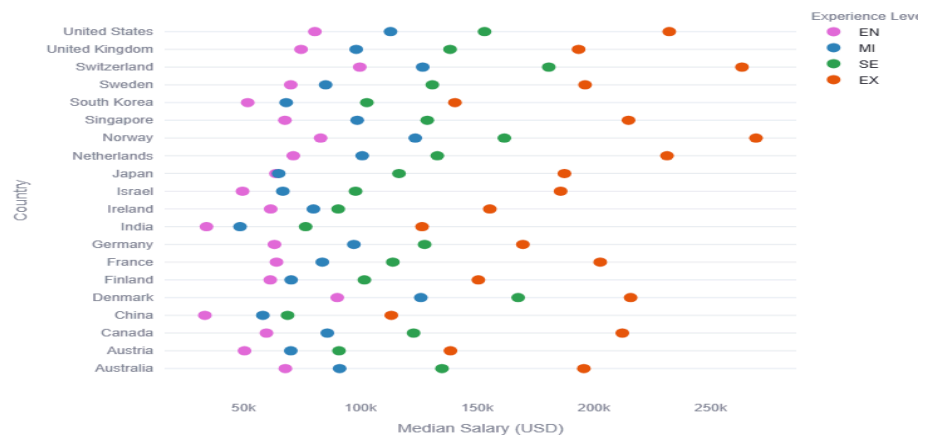
ניתוח החלופה השנייה:

ב. סקיצה של החלופה:

1. Boxplot – Salary Distribution by Job Title



2. Dot Plot – Median Salary by Country and Experience Level



ג. תיאור החלופה במונחים של Marks & Channels:

Graph 1: Boxplot – Salary Distribution by Job Title

Marks:

- Area marks boxes - כל תיבה מייצגת את ה-25%–75% IQR של השכר עבור תפקיד מסוים. זה mark שמסייע להראות שונות פנימית ולא רק "מספר אחד" לכל תפקיד.
- Line marks median + whiskers - הקו בתוך התיבה הוא החציון, וה-whisker מציגים טווח טיפוס. בכך מתקבל סיכום חזק של ההתפלגות שמאפשר להשוות "מרכז" מול "פיזור" בין תפקידים.
- Point marks outliers - נקודות חריגים מצביעות על משכורות נדירות/קיצוניות שאינן חלק מהטווח הטיפוסי. זה תומך במשימה של איתור חריגים (locate) בלי להסתיר אותם בתוך הסיכום.

Channels:

- Position (x) – Job Title (categorical, unordered) המיקום האופקי מבדיל בין קטגוריות תפקיד. מאחר שאין סדר טבעי לתפקידים, המיון שביצעת לפי חציון הוא החלטת עיצוב שמשרתת השוואה.
- Position (y) – Salary (quantitative) ציר ה-y מאפשר השוואה מדויקת של ערכי שכר בין תפקידים, משום שמיקום הוא הערוץ התפיסתי המדויק ביותר להשוואה כמותית.
- Color (single hue) – neutral / no additional encoding כאן צבע לא מקודד משתנה נוסף, אלא משמש כאחידות ויזואלית. זה מפחית עומס ומונע פרשנות שגויה ("אם צבע שונה – יש משמעות").

Graph 2: Dot Plot – Median Salary by Country and Experience Level

:Marks

- Point marks - כל נקודה מייצגת חציון שכר עבור צירוף של מדינה ורמת ניסיון. זה mark "דחוס" שמאפשר להציג הרבה קבוצות בלי להעמיס כמו boxplot לכל קבוצה.

:Channels

- Position (x) – Median Salary (quantitative)
שוב נבחר המיקום, כי המטרה היא השוואות בין חציונים: מי גבוה יותר, כמה פער יש, ועוד. זה הערוץ המדויק ביותר להשוואות כמותיות, ולכן מתאים מאוד למשימת compare.
- (Position (y) – Country (categorical)
המיקום האנכי מפריד בין מדינות ומאפשר להשוות בתוך אותה מדינה לאורך רמות ניסיון שונות. בנוסף, הוא מאפשר להשוות מדינות שונות באותה רמת ניסיון על ידי "סריקה" אופקית.
- Color hue – Experience Level (categorical, conceptually ordered)
צבע בגוון (hue) מתאים לקטגוריות, כי המטרה היא להבדיל בין EN/MI/SE/EX ולא למדוד גודל כמותי. למרות שיש סדר מושגי בין הרמות, ההשוואה הכמותית עדיין נעשית באמצעות x ולא באמצעות הצבע וזה מונע הטעיה.
- Size – constant
הגודל קבוע כדי לא "להמציא" משמעות שלא קיימת בנתונים. זה מצמצם סיכון של פירוש שגוי (למשל לחשוב שגודל = מספר משרות).

ד. אקספרסיביות ואפקטיביות

אקספרסיביות:

- כל ערוץ מייצג רק מידע שקיים בנתונים, אין שימוש בצבע/גודל/צורה כדי "לרמוז" על משתנה שלא מופיע בטבלה. כך נמנעת הטעיה סמנטית.
- אין קידוד כפול של אותו משתנה, השכר מוצג כמותית באמצעות מיקום הצבע ב-dot plot מקודד רק ניסיון ולא את השכר שוב.
- Boxplot - צבע ניטרלי כדי לא ליצור משמעות נוספת, שימוש בגוון אחיד מונע מצב שבו המשתמש חושב שהצבע מייצג קטגוריה/איכות/קבוצה נוספת שלא קיימת.
- Dot plot - צבע מייצג משתנה קטגורי (experience level), זה מיפוי "נקי". אין ניסיון להשתמש בצבע כדי למדוד "כמה" אלא כדי להבדיל בין קבוצות.
- הימנעות מערצים שעלולים להטעות, לא נעשה שימוש בשטח/נפח/גדלים משתנים ללא משמעות, ולכן לא נוצרת היררכיה חזותית מזויפת.

אפקטיביות:

- Position לערכים כמותיים = השוואה מדויקת יותר, בשני הגרפים השכר ממופה למיקום, ולכן קל לבצע compare בין תפקידים/מדינות/רמות ניסיון בצורה מדויקת יחסית.

- Color משמש להפרדה קטגוריאלית ולא למדידה, ב-Dot plot הצבע נועד לזהות את קבוצת הניסיון. ההשוואה של "מי גבוה יותר" נשענת על מיקום, מה שמעלה דיוק תפיסתי.
- פיצול לשני גרפים תומך במיקוד מטלות, במקום "לדחוף הכל" לגרף אחד, כל גרף ממקסם אפקטיביות למטרה אחרת.
- Boxplot אפקטיבי במיוחד ל-distribution + outliers, הוא מציג גם מרכז (median) וגם פיזור (IQR) וגם חריגים. זה תומך במטלות locate ו-compare של התפלגויות.
- Dot plot אפקטיבי ל-summary comparisons בקנה מידה גדול, הוא מאפשר להשוות הרבה מדינות ורמות ניסיון בזמנית, תוך שמירה על גרף "דחוס" שלא מתפוצץ בעומס marks.
- הצדקה לאי שימוש במספרים על הגרף, המטרה היא overview/compare/summarize ולא lookup מדויק. מספרים בכל תא/נקודה היו מוסיפים עומס ויזואלי בלי לשרת את המשימה המרכזית.

ה. יתרונות וחסרונות:

יתרונות:

- שילוב בין Distribution view ל-Summary view, ה-boxplot מסביר את "מה קורה בתוך התפקיד" (שונות, חריגים), וה-dot plot מסביר "איך זה משתנה בין מדינות וניסיון" בחציון.
- כיסוי טוב של מגוון מטלות משתמש, מתקבל מענה ל-compare בין תפקידים / מדינות / ניסיון, summarize דפוסים כלליים, identify patterns פערים עקביים, locate outliers ב-boxplot.
- שימוש בערוצים בעלי רזולוציה תפיסתית גבוהה, הצגת שכר באמצעות מיקום מאפשרת להבחין בפערים בצורה טובה יותר מאשר ערוצים פחות מדויקים כמו צבע או שטח.

חסרונות:

- Boxplot פחות נוח להשוואה מספרית "מדויקת" של חציונים, אפשר לראות מי גבוה יותר, אבל קשה להוציא ערך מדויק בלי hover/טבלה - זה trade-off לטובת overview
- Dot plot לא מציג פיזור או שונות פנימית, הוא מראה רק חציון, ולכן שני מקומות עם אותו חציון יכולים להיות בעלי התפלגות שונה לגמרי - מידע שאובד.
- ריבוי קטגוריות (מדינות) דורש יותר קשב/גלילה, כשיש הרבה מדינות, המשתמש צריך לסרוק לאורך ציר y זה מגדיל מאמץ תפיסתי בהשוואות רחבות.
- תלות בבחירת צבעים ב-dot plot, אם הגוונים קרובים מדי או לא מותאמים לעיוורון צבעים, ההבחנה בין קטגוריות נפגעת. לכן חייבים palette מובחן ויציב.

2. הצדקת הבחירה בחלופה הסופית:

החלופה השנייה נבחרה למימוש הסופי משום שהיא מספקת מענה רחב ומדויק יותר למטלות שהוגדרו, תוך יישום מאוזן של עקרונות האקספרסיביות והאפקטיביות של Munzner. השילוב בין boxplot להתפלגות שכר לפי תפקיד לבין dot plot המציג שכר חציוני לפי מדינה ורמת ניסיון מאפשר ניתוח רב-ממדי של הנתונים מבלי לדחוס משתנים רבים לגרף יחיד. חלוקה זו תומכת הן במטלות של השוואה וזיהוי מגמות והן בזיהוי שונות פנימית וערכים חריגים, שאינם ניתנים לזיהוי ברור בחלופה הראשונה. בנוסף, השימוש בערוצים ויזואליים אפקטיביים - מיקום להשוואות כמותיות וצבע להבחנה קטגוריאלית - משפר את הדיוק התפיסתי ומפחית עומס קוגניטיבי. לפיכך, החלופה השנייה

מאפשרת עומק אנליטי גבוה יותר לצד קריאות חזותית טובה, ומשקפת התאמה מובהקת יותר למטרות הניתוח ולמודל התכנון של Munzner.

חלק 4: הסבר על הייצוגים ויישומם:

לפני התחלת העבודה עם המידע נדרש ביצוע Pre-processing ראשוני על כלל הדאטה שכלל חיבור שני קבצי המידע. עשינו זאת באופן הבא: המידע במקור הגיע מאתר Kaggle בשני קבצים שונים ולכן נאלצנו לאחד אותם לקובץ אחד. עשינו זאת ראשית במחיקת עמודת "Salary_Local" אשר קיימת רק בקובץ השני ולא תרמה לשאלת המחקר, ולאחר מכן איחוד הדאטאות תוך תיקון שדה "job_id" בקובץ שהעברנו כך ש $job_id_new = job_id_old + 15,000$ כדי לאפשר מזהה ייחודי לכל רשומה. נציין כי ע"מ להריץ את הדאשבורד יש להוריד את העדכון החינמי של streamlit לגרסה 1.35.0, עדכון זה

הינו חינמי

מטלה 1: האם קיימים הבדלים גיאוגרפיים מובהקים באימוץ AI ברחבי העולם?

הפרויקט תוכנן לאפשר למשתמש לחקור את שוק ה-AI משתי פרספקטיבות משלימות: צד ה"היצע" (העובדים) וצד ה"ביקוש" (החברות). כדי למנוע עומס קוגניטיבי, הפרדנו את המידע לשלושה מבטים (Views) נפרדים, ביניהם ניתן לעבור בלחיצת כפתור, תוך שמירה על עקביות ויזואלית ופונקציונאלית.

סך 1: Comparator

א. עיבוד מוקדם של הנתונים (Pre-processing) בשלב טעינת הנתונים מתוך קובץ ה-CSV, ביצענו מספר פעולות עיבוד חיוניות כדי לאפשר את הניתוח הגיאוגרפי והכלכלי המוצג בדאשבורד:

- הנדסת מאפיינים - זיהוי מקור עובדים: יצרנו משתנה בוליאני חדש בשם is_same_location. משתנה זה מחושב על ידי השוואה בין עמודת employee_residence (מגורי העובד) לבין company_location (מיקום החברה). פעולה זו מאפשרת לנו לפלח בזמן אמת את כוח העבודה ל"מקומי" (מתוך המדינה) מול "בינלאומי/מיקור חוץ" (מחוץ למדינה), נתון קריטי להבנת דפוסי העסקה גלובליים.
- אגרגציה לטיפול בערכים חסרים: כדי ליצור את גרף "סולם השכר" (Salary Comparison) ולהשוות אותו לממוצע העולמי, חישבנו מראש את ממוצעי השכר הגלובליים לפי שנות ניסיון (global_exp_avg). חישוב מקדים זה מאפשר לנו להציג קו ייחוס ירוק ("Global Avg") בכל תרשים, גם כאשר בוחרים מדינה שיש לה נתונים חלקיים בלבד עבור שנות ניסיון מסוימות.
- המרה וטיוב: ביצענו המרה של נתונים בוליאניים למספרים (Integer) כדי לאפשר ספירה וחישוב אחוזים מדויק עבור תרשים הבר האופקי המציג את התפלגות מקור העובדים.

ב. שינויים במימוש הוויזואליזציה לעומת התכנון המקורי במהלך הפיתוח, נדרשנו לבצע התאמות כדי לשפר את חווית המשתמש בדאשבורד הראשי:

1. שיפור הקוד הוויזואלי:

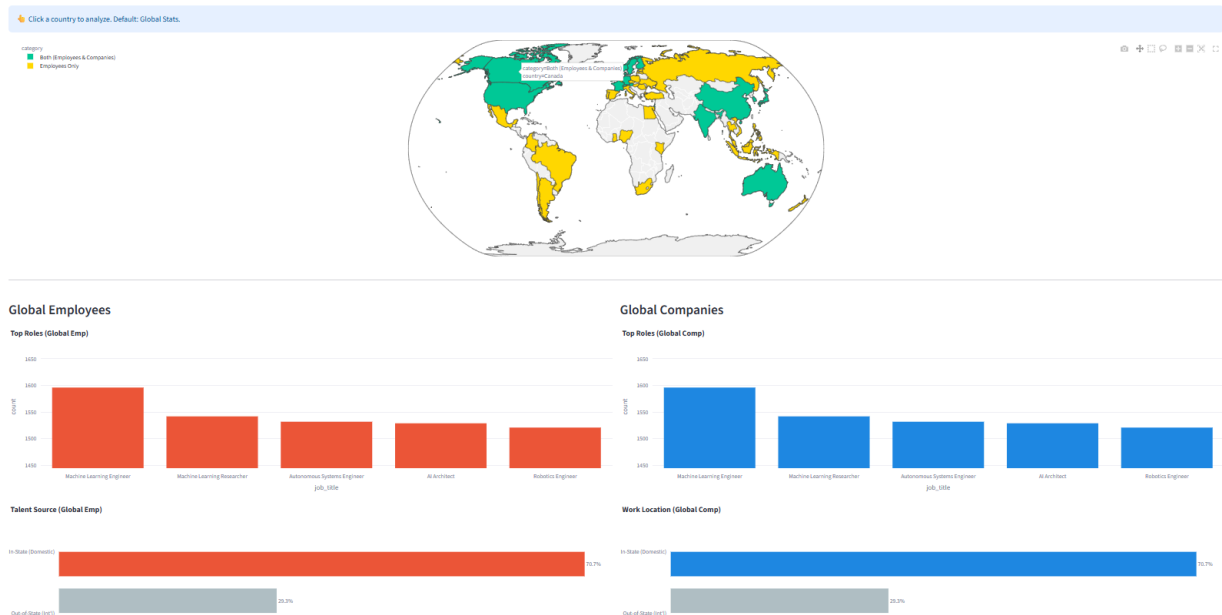
- קנה מידה דינמי בציר ה-Y: בתכנון המקורי, גרף התפקידים המובילים ("Top Roles") התחיל תמיד מאפס. נוכחנו לדעת כי במדינות מסוימות, הפערים בין מספר המשרות בתפקידים השונים הם קטנים, והגרף נראה "שטוח". כפתרון, הטמענו לוגיקה בקוד (fig1.update_yaxes) המזהה

את הערך המינימלי בגרף ומגדירה את תחילת הציר ב-95% מערך זה. שינוי זה מדגיש את ההבדלים הוויזואליים גם בפערים קטנים.

- מנגנון השוואה (Comparator Mode): במקור תכננו להציג רק נתונים עבור מדינה בודדת. תוך כדי מימוש, הבנו שהערך האמיתי נוצר מהשוואה. לכן, פיתחנו את מצב "Comparator", המשתמש במפה לא רק לבחירה בודדת אלא לבחירה כפולה (מדינה ראשית ומשנית), ומחליף את התצוגה הגלובלית בתצוגה מפוצלת המעמתת נתונים של שתי מדינות זו לצד זו.
- כפתור Global Countries: שילבנו כפתור ייעודי בראש הממשק המאפשר גישה מהירה לנתוני מאקרו. במבט זה, לחיצה על הכפתור חושפת טבלה המדרגת את 10 המדינות המובילות במספר העובדים ומספר מקומות העבודה. פונקציה זו מאפשרת למשתמש לזהות במהירות את מוקדי הכוח של הטאלנט העולמי מבלי לסרוק את המפה כולה, בנוסף במצב דיפולטי אנו מציגים את הנתונים של מדינות אלה כנתונים גלובליים.

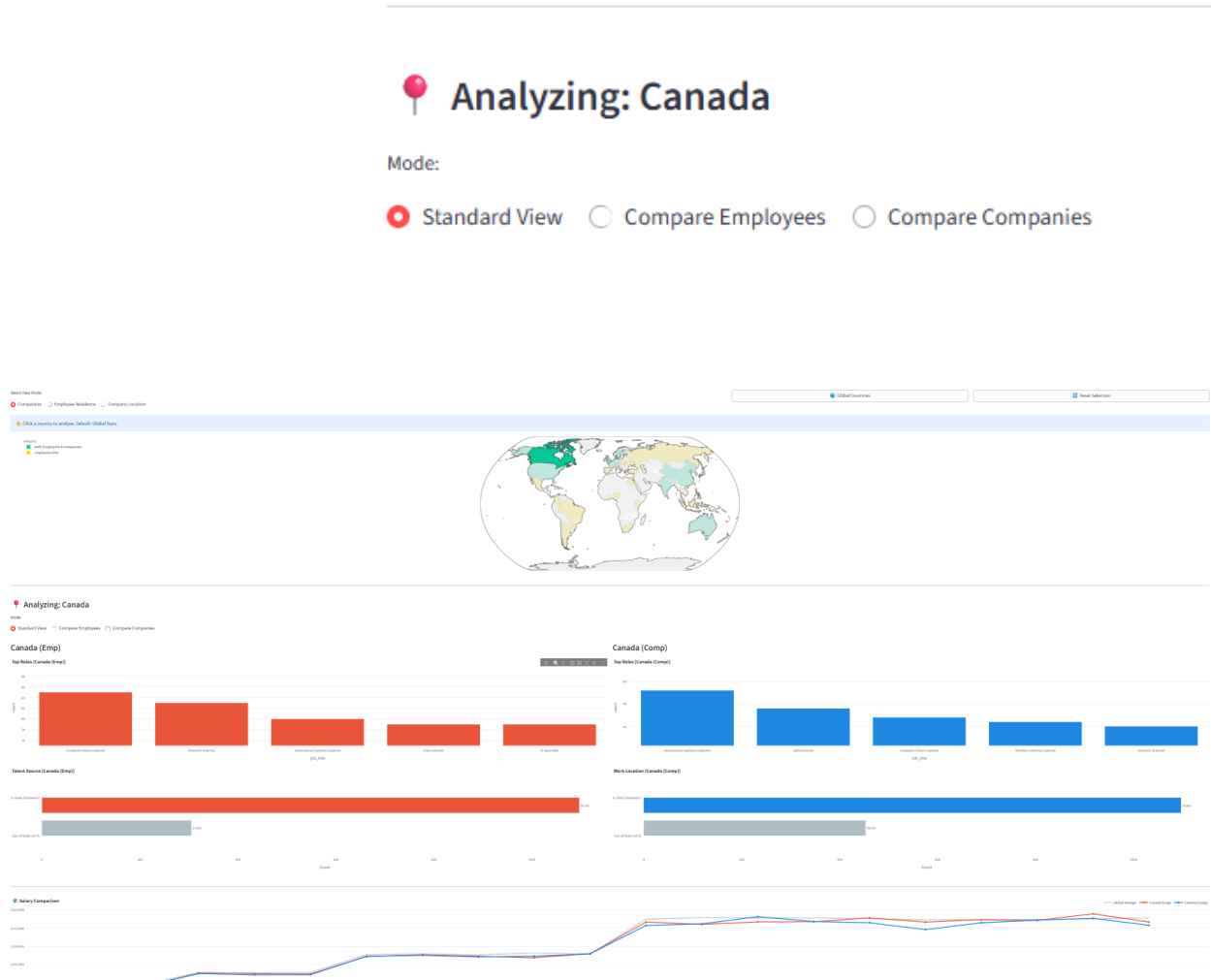
2. ניהול מצב (State Management): המעבר בין תצוגת "מגורי עובדים" לתצוגת "מיקום חברות" דרש שינוי ארכיטקטוני. במקום לטעון מחדש את כל הדף ולאבד את המדינה שנבחרה, הטמענו שימוש ב-st.session_state כדי "לזכור" את המדינה הראשית והמשנית גם בעת החלפת מצבי תצוגה או רענון רכיבים.

ג. אופן המימוש
תמונה 1:

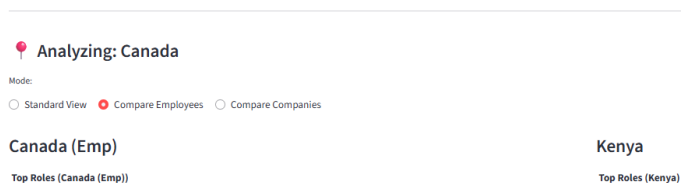


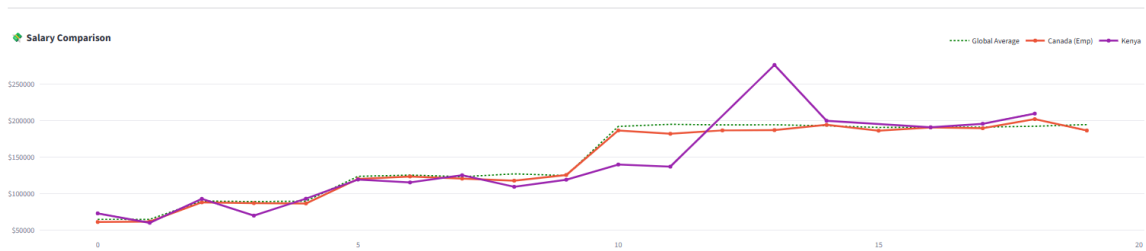
מסך הבית - מבט גלובלי (Global Landscape) תצלום זה מציג את מצב ברירת המחדל. המפה האינטראקטיבית צבועה לפי סוג הפעילות (עובדים/חברות/גם וגם), ומתחתיה מוצגים ארבעה גרפים (תפקידים, מקור עובדים, שכן לפי ניסיון, וסולם שכר) עבור כלל העולם.

תמונה 2: אינטראקציה - Drill-down למדינה (Single Country View) כאשר המשתמש לוחץ על מדינה במפה (למשל, קנדה), הקוד מזהה את אירוע הבחירה (on_select). הכותרות משתנות ל-"Analyzing: Canada", וארבעת הגרפים מתעדכנים מיידית להציג רק את הנתונים הרלוונטיים למדינה זו, תוך שמירה על קו הייחוס הגלובלי (הקו הירוק המקווקו) להשוואה.



תמונה 3: מצב השוואה (Comparator Mode) במצב זה, המשתמש בוחר שתי מדינות שונות במפה. המסך מתפצל לשתי עמודות: עמודה ימנית למדינה הראשית ועמודה שמאלית למדינה המשנית. בתחתית המסך נוצר גרף קווי ייעודי (create_comparison_line_chart) המציג את עקומות השכר של שתי המדינות שנבחרו על גבי מערכת צירים אחת, ומאפשר זיהוי מהיר של פערים כלכליים.





ד. כתיבת הקוד: בקוד שנכתב עבור וויזואליזציה זו נעשה שימוש בספריות ובכלים הבאים:

- Streamlit: עבור בניית ממשק ה-Frontend (צד לקוח) וניהול לוגיקת האפליקציה. נעשה שימוש נרחב ב-st.session_state לצורך ניהול מורכב של בחירת מדינות (Primary ו-Secondary) ולמניעת איפוס הבחירה בעת שינוי מסננים או מעבר בין מצבי תצוגה.
- Pandas: לטובת עיבוד וניהול הנתונים, החל מטעינת המידע ועד להכנתו להצגה ויזואלית. השימוש בספרייה אפשר להעביר את ה-DataFrame כקלט לפונקציות גנריות המייצרות את הגרפים.
- Plotly Express & Graph Objects: עבור יצירת הוויזואליזציות האינטראקטיביות. הקוד נבנה בצורה מודולרית סביב פונקציית create_graphs, גישה שאפשרה שימוש חוזר באותו קוד בדיוק – הן עבור התצוגה הגלובלית והן עבור התצוגות המפוצלות במצב ההשוואה.

מסך 2: מגורי עובדים (Employee Residence View)

- א. עיבוד מוקדם של הנתונים (Pre-processing) במבט זה, המיקוד הוא בצד ה"היצע" של שוק העבודה. ביצענו עיבוד ייעודי לנתונים כדי לזהות דפוסי הגירה דיגיטלית:
1. פילוח גאוגרפי: הנתונים עברו קיבוץ (Grouping) לפי עמודת employee_residence. פעולה זו מאפשרת לנו להתייחס למיקום הפיזי של העובד כנקודת המוצא לניתוח, ללא קשר למיקום החברה המעסיקה.
 2. הנדסת מאפיינים - ניתוח מקור התעסוקה: השימוש במשתנה is_same_location מקבל כאן משמעות ייחודית. במבט זה, הנתון משמש להצגת אחוז העובדים במדינה מסוימת המועסקים על ידי חברות זרות (מיקור חוץ/Remote) לעומת חברות מקומיות, מה שמעיד על איכות הטאלנט המקומי והביקוש הגלובלי אליו.
- ב. שינויים במימוש הוויזואליזציה לעומת התכנון המקורי במהלך הפיתוח ביצענו התאמות כדי להבדיל ויזואלית בין מסך זה למסך החברות:

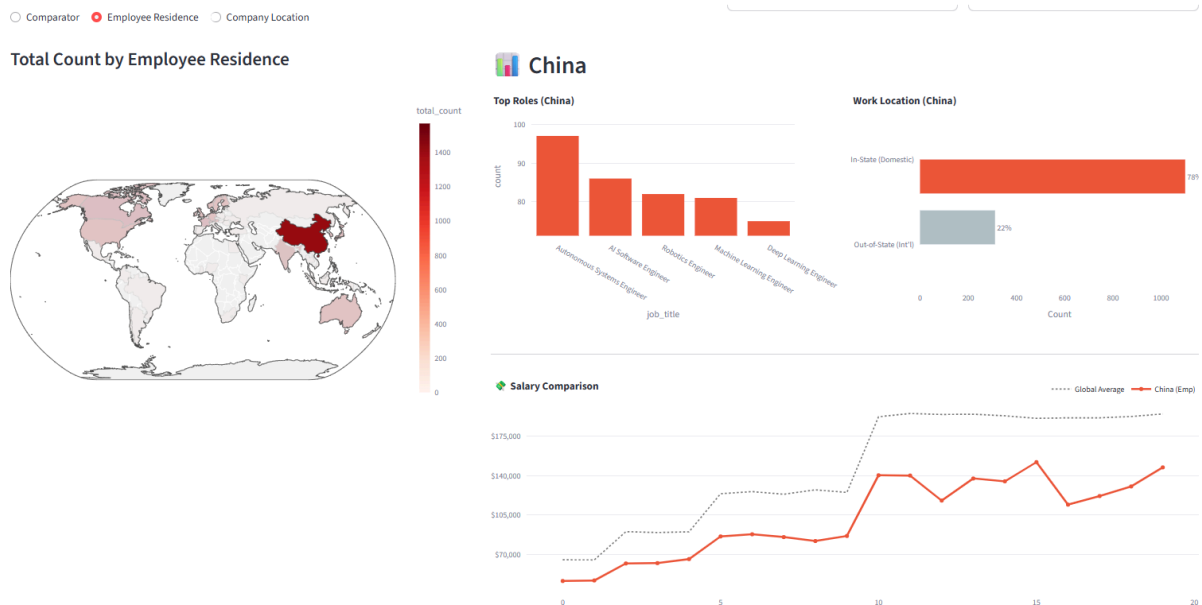
1. שפה חזותית: המבט מאופיין בשימוש עקבי בצבע אדום. בחירה זו מאפשרת למשתמש להבין אינטואיטיבית שהוא צופה בנתוני עובדים, ומונעת בלבול במעבר בין המסכים.
2. המפה: מציגה מפת חום (Choropleth) של ריכוזי עובדים. סקאלת הצבעים מנורמלת ביחס למקסימום הגלובלי (GLOBAL_MAX_COUNT), כך שמדינות עם ריכוז גבוה של מומחי AI בולטות בכהותן באופן פרופורציונלי.

3. כפתור Global Countries: שילבנו כפתור ייעודי בראש הממשק המאפשר גישה מהירה לנתוני מאקרו. במבט זה, לחיצה על הכפתור חושפת טבלה המדרגת את 10 המדינות המובילות במספר העובדים. פונקציה זו מאפשרת למשתמש לזהות במהירות את מוקדי הכוח של הטאלנט העולמי מבלי לסרוק את המפה כולה. בנוסף, במצב דיפולטי אנו מציגים את הנתונים של מדינות אלה כנתונים גלובליים.
4. לוח המחוונים (Dashboard): בחירה במדינה במפה מפעילה סינון רוחבי על שלושה תרשימים:

- Top Roles: מציג את התפקידים הנפוצים ביותר בקרב התושבים (למשל, Data Scientists).
- Work Location: גרף ברים אופקי הממחיש כמה מהתושבים עובדים "פנימה" (לשוק המקומי) וכמה "החוצה" (לחברות בינלאומיות).
- Salary Comparison: משווה את עקומת השכר של התושבים לממוצע הגלובלי (קו ירוק מקווקו).
- Talent Source: משווה בין כמה אחוז מהעובדים עובדים בחברות במדינה הנבחרת לעומת כאלה שעובדים במדינות אחרות מהרשימה

ג. אופן המימוש

תמונה 1: מסך Employee Residence בתצלום זה המשתמש בחר בסין. הגרפים הכחולים מציגים את התפלגות התפקידים של תושבי סין ואת שכרם הממוצע בהשוואה לעולם.



מסך 3: מיקום חברות (Company Location View)

א. שינויים בPreprocessing למרות שהמבנה הכללי זהה למסך הקודם (לצורך עקביות), הלוגיקה והתוכן משתנים כדי לשקף את צד ה"חברות":

1. מקור הנתונים: הסינון והאגרגציה מתבצעים הפעם לפי עמודת company_location.
 2. משמעות המדדים: הגרף המציג "מקומי מול זר" מקבל משמעות הפוכה – הוא מראה כמה מכוח העבודה של החברות במדינה זו הוא מקומי, וכמה גויס ממדינות אחרות (Offshore/Remote).
 3. ציר השכר: מייצג את "נכונות לשלם" (Willingness to Pay) של החברות באזור זה, ולא בהכרח את יוקר המחיה של העובדים.
- ב. שינויים במימוש הוויזואליזציה לעומת התכנון המקורי ההתאמות העיקריות נועדו לחדד את ההבדלים בין מדינות עם שווקים צפופים:

1. שפה חזותית: המבט מאופיין בשימוש עקבי בצבע כחול שינוי הצבע הוא קריטי למניעת עומס קוגניטיבי בעת המעבר בין המצבים.
 2. דינמיות בצירים: בדומה למבט הראשון, גם כאן יושם מנגנון לחישוב דינמי של ציר ה-Y בגרפים (* min 95.0). מנגנון זה פותח כדי להבליט פערים קטנים בין פופולריות של משרות בחברות השונות, שלא היו נראים בגרף המתחיל מאפס.
 3. כפתור Global Countries: הכפתור משנה את תפקודו בהתאם למבט. במבט זה, לחיצה על הכפתור חושפת טבלה המדרגת את 10 המדינות המובילות במספר המשרות (החברות). פונקציה זו מאפשרת למשתמש לזהות היכן מרוכז ההון והיכן נמצאים ה-Tech Hubs הגדולים. בנוסף, במצב דיפולטי אנו מציגים את הנתונים של מדינות אלה כבסיס להשוואה.
- ג. אופן המימוש
- תמונה 1: מבט Company Location בתצלום זה המשתמש בחר באוסטרליה. ניתן לראות כיצד מפת העולם והגרפים (בצבע כחול) משקפים כעת את הנתונים מנקודת המבט של המעסיקים האוסטרליים, כולל נכונותם לשלם עבור טאלנטים בינלאומיים.



ד. כתיבת הקוד: (משותף למסכים 2 ו-3)

בקוד שנכתב עבור שני המבטים הללו, נעשה שימוש בספריות ובכלים הבאים:

- Streamlit: שימשה לבניית ממשק המשתמש ולניהול הניווט בין המבטים (באמצעות רכיב `st.radio`).
- Pandas: שימשה לטעינת קובץ ה-CSV ולביצוע מניפולציות על הנתונים (כגון Grouping ו-Filtering) בזמן אמת בהתאם למבט שנבחר (`view_mode`).
- Plotly Express & Graph Objects: שימשו ליצירת הוויזואליזציות האינטראקטיביות. הבחירה ב-Plotly אפשרה לנו להטמיע יכולות Zoom ו-Hover מתקדמות בשני המבטים.
- שימוש חוזר בקוד (Code Reusability): כדי לשמור על קוד נקי ויעיל, נמנענו משכפול קוד. במקום זאת, כתבנו פונקציה גנרית אחת בשם `create_graphs`. פונקציה זו מקבלת כפרמטרים את ה-DataFrame המסונן, הכותרת הרצויה וצבע הנושא (אדום או כחול). כך, החלפת מבט מעדכנת את כל הממשק ללא טעינת דף מחדש.
- ניהול מצב (State Management): אתגר מרכזי היה לשמור על בחירת המדינה בעת המעבר בין המבטים. השתמשנו ב-`st.session_state` כדי לשמור את המשתנה `primary_country`. בזכות כך, האינטראקציה היא רציפה: המשתמש יכול לבחור מדינה במבט אחד, ולעבור למבט השני כדי לראות מיד את הנתונים המשלימים עבור אותה מדינה.
- שימוש בכלי עזר (AI): במהלך הפיתוח נעשה שימוש במודלי שפה לצורך אופטימיזציה של שאלות Pandas, כתיבת הלוגיקה לחישוב הטווח הדינמי של ציר ה-Y, ופתרון בעיות תאימות (Debugging) הקשורות לאירוע הבחירה (`on_select`) בגרסאות חדשות של Streamlit.

מטלה 2: ניתוח השוואתי רב מימדי של תפקידי עבודה בתחום ה-AI.

עיבוד מקדים נוסף: בשלב העיבוד המוקדם הנתונים נטענו מקובץ CSV ועברו אגרגציה לפי `job_title`, כך שכל תפקיד מיוצג בערך אחד לכל מדד. עבור מדדים כמותיים (כגון שכר, ניסיון, יחס עבודה מרחוק והטבות) חושב ערך ממוצע, ועבור מדד הביקוש חושבה ספירה של מספר המשרות.

מאחר שהמדדים נמדדים בסקאלות שונות, בוצע נירמול מסוג Min-Max לכל מדד בנפרד לטווח $[0,1]$, על מנת לאפשר השוואה ויזואלית בין מדדים הטרוגניים. במקביל נשמרו הערכים הגולמיים, אשר מוצגים באינטראקציות hover ובטבלת נתונים משלימה. בנוסף, הוחל ערך מינימלי ויזואלי לצורכי תצוגה, המאפשר אינטראקציה גם עבור ערכים מנורמלים נמוכים במיוחד.

תצוגה דפולטית:

Multidimensional Comparison of AI Job Roles

Compare AI job roles across disparate metrics (Salary, Frequency, Remote availability) using Min-Max Normalization.

Select Metrics for Comparison

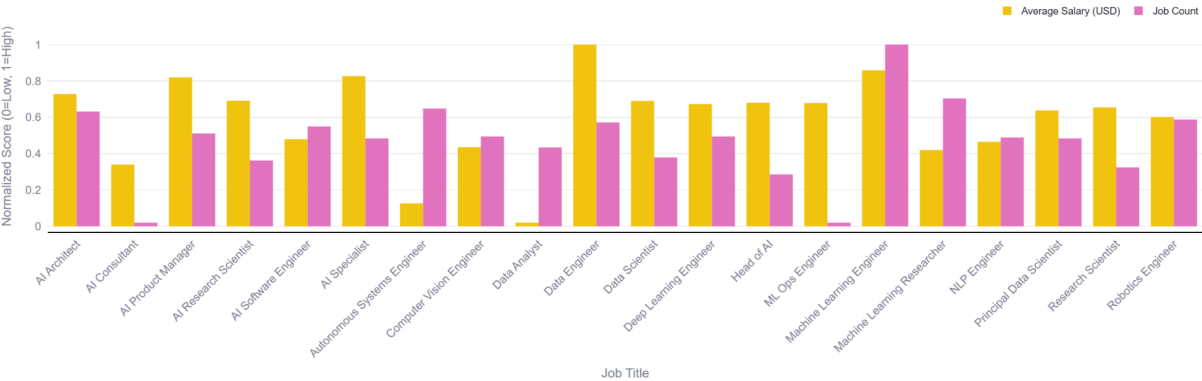
- ☒ Average Salary (USD)
- ☒ Job Count
- ☐ Years of Experience
- ☐ Benefits Score
- ☐ Remote Work Ratio (%)

Chart Sort Settings

Sort Chart By:

Name (A-Z)

Normalized Multidimensional Comparison (Sorted by: Name (A-Z))



Underlying Data (Raw Values)

Search by Job Title

Type job name...

job_title	Average Salary (USD)	Job Count
AI Architect	119,725.43	1,529.00
AI Consultant	116,733.32	1,417.00
AI Product Manager	120,434.36	1,507.00
AI Research Scientist	119,444.34	1,480.00
AI Software Engineer	117,807.11	1,514.00
AI Specialist	120,487.82	1,502.00
Autonomous Systems Engineer	115,083.59	1,532.00
Computer Vision Engineer	117,467.19	1,504.00
Data Analyst	114,109.41	1,493.00
Data Engineer	121,828.42	1,518.00
Data Scientist	119,434.51	1,483.00
Deep Learning Engineer	119,301.77	1,504.00
Head of AI	119,360.61	1,466.00
ML Ops Engineer	119,346.68	1,414.00

כדי לבחור עוד קריטריונים להשוואה ניתן לסמן אותם בסרגל Select Metrics for Comparison

ניתן לבחור בין אחד לחמישה פיצ'רים שאותם אנחנו רוצים להשוואות ולראות את השוני ביניהם

Select Metrics for Comparison

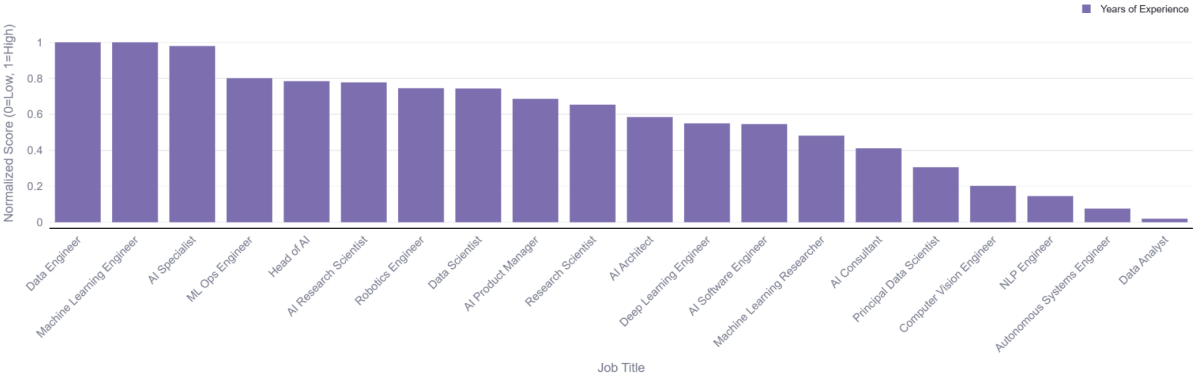
☐ Average Salary (USD) ☐ Job Count ☒ Years of Experience ☐ Benefits Score ☐ Remote Work Ratio (%)

Chart Sort Settings

Sort Chart By:

Years of Experience

Normalized Multidimensional Comparison (Sorted by: Years of Experience)



Select Metrics for Comparison

☒ Average Salary (USD) ☒ Job Count ☒ Years of Experience ☒ Benefits Score ☒ Remote Work Ratio (%)

Chart Sort Settings

Sort Chart By:

Name (A-Z)

Normalized Multidimensional Comparison (Sorted by: Name (A-Z))



בנוסף ניתן לעמוד על עמודה רצויה ולקבל את הערך המקורי שלה ואת הערך היחסי לאחר הנרמול
וכן למיין את סדר העבודות לפי פיצר מסוים לבחירה

Select Metrics for Comparison

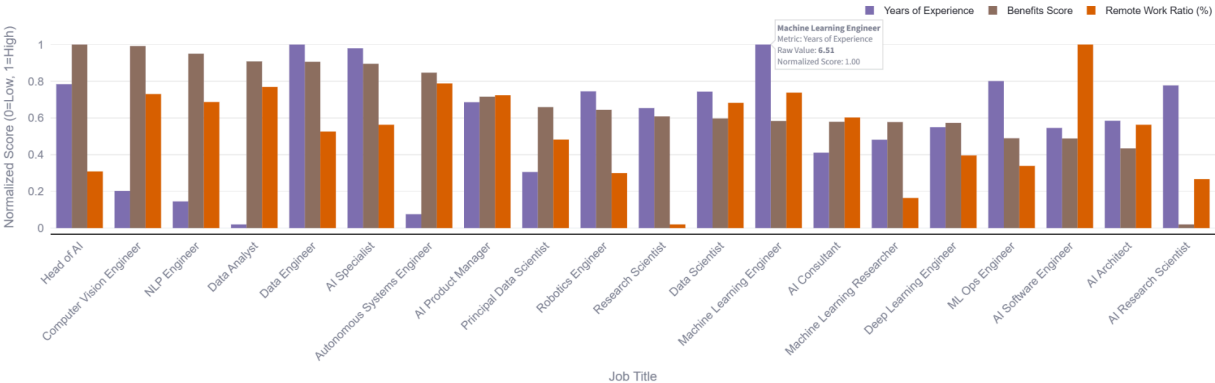
☐ Average Salary (USD) ☐ Job Count ☒ Years of Experience ☒ Benefits Score ☒ Remote Work Ratio (%)

Chart Sort Settings

Sort Chart By:

Benefits Score

Normalized Multidimensional Comparison (Sorted by: Benefits Score)



בנוסף יש טבלה עם מפת חום שבה כל שורה מייצגת עבודה אחרת והעמודות נקבעות על פי הפיצרים שנבחרו למעלה. הערכים בטבלה הם הערכים הממוצעים של כל שילוב ולכל עמודה יש מפת חום - ככל שהצבע יותר כהה ככה הערך יותר גדול

Underlying Data (Raw Values)

Search by Job Title

Type job name...

job_title	Benefits Score	Remote Work Ratio (%)	Years of Experience
AI Architect	7.47	50.00	6.31
AI Consultant	7.49	50.21	6.23
AI Product Manager	7.51	50.86	6.36
AI Research Scientist	7.42	48.41	6.41
AI Software Engineer	7.48	52.34	6.30
AI Specialist	7.53	50.00	6.50
Autonomous Systems Engineer	7.52	51.21	6.07
Computer Vision Engineer	7.54	50.90	6.13
Data Analyst	7.53	51.11	6.03
Data Engineer	7.53	49.80	6.51

בטבלה מוצגים כל סוגי העבודות והעמודות מייצגות את הפיצרים שנבחרו למעלה

ניתן לחפש עבודות ספציפיות על פי שמם
לדוג שם עבודה עם AI (השמות של סוגי העבודות השונות הם כמו ציר X בגרף העמודות)

Underlying Data (Raw Values)

Search by Job Title

AI

job_title	Benefits Score	Remote Work Ratio (%)	Years of Experience
AI Architect	7.47	50.00	6.31
AI Consultant	7.49	50.21	6.23
AI Product Manager	7.51	50.86	6.36
AI Research Scientist	7.42	48.41	6.41
AI Software Engineer	7.48	52.34	6.30
AI Specialist	7.53	50.00	6.50
Head of AI	7.54	48.64	6.41

ספריות בהן נעשה שימוש:

- Pandas - לטעינת הנתונים, עיבוד ואגרגציה לפי תפקיד עבודה (חישוב ממוצעים וספירות).
- Streamlit - לבניית ממשק משתמש אינטראקטיבי הכולל בחירת מדדים, הצגת גרפים וטבלאות.
- Plotly Express - ליצירת גרפי עמודות אינטראקטיביים עם תמיכה ב-hover והשוואה בין מדדים.

הקוד ממומש כאפליקציית Streamlit ומחולק למספר שלבים לוגיים ברורים. תחילה מוגדר מבנה העמוד והכותרות, ולאחר מכן נטען הקובץ ומבוצעת בדיקת תקינות לקיומו.

ממשק המשתמש מאפשר בחירה דינמית של מדדים באמצעות תיבות סימון, כאשר כל שינוי בבחירה גורם לרענון מיידי של הוויזואליזציה. בחירה זו משפיעה הן על הנתונים המוצגים והן על עיצוב הגרף (כגון מספר העמודות, צבען והמקרא).

הגרף עצמו ממומש כתרשים עמודות מקובץ (Grouped Bar Chart), כאשר לכל מדד מוקצה צבע קבוע ועקבי. הערכים המוצגים בגרף הם ערכים מנורמלים, אך בעת מעבר עם העכבר על כל עמודה מוצג חלון מידע (hover) הכולל את שם התפקיד, שם המדד, הערך הגולמי והערך המנורמל.

בנוסף לגרף, הקוד מציג טבלת נתונים משלימה המציגה את הערכים הגולמיים של כל המדדים לכל תפקיד, עם קידוד צבעים רציף. טבלה זו מאפשרת אימות של המידע המוצג בגרף והעמקה בניתוח, ללא צורך באינטראקציה עם התרשים עצמו.

לבסוף, הוגדרו מאפייני עיצוב כלליים (רקע לבן, סידור מקרא, סיבוב תוויות ציר) על מנת לשפר את קריאות הוויזואליזציה ולהתאימה להצגה בדו"ח או במצגת.

מטלה 3: השוואה בין רמת הביקוש למיומנויות שונות בשוק העבודה בתחום ה-AI לבין רמת השכר המוצעת עבורן.

א. ביצוע Pre-processing:

במסגרת יצירת הוויזואליזציה בוצע תהליך עיבוד מקדים על בסיס הנתונים הגולמיים כדי להתאימם לניתוח סטטיסטי ולהצגה ויזואלית זו. הנתונים המקוריים הגיעו ברמת "משרה בודדת", בעוד שמטרת הניתוח היא בחינת "מיומנויות טכנולוגיות".
להלן שלבי העיבוד שבוצעו:

1. פירוק מחרוזות כישורים: המידע בעמודת required_skills הופיע במקור כמחרוזת טקסט אחת המכילה מספר כישורים מופרדים בפסיקים (למשל: "Python, SQL, AWS"). בוצעה פעולת המרה שהפכה כל מחרוזת כזו לרשימה של כישורים נפרדים.
2. שינוי מבנה הנתונים: בוצעה פעולת explode על רשימות הכישורים. פעולה זו שינתה את יחידת הניתוח הבסיסית של הטבלה: משורה המייצגת משרה, לשורות המייצגות כישור בודד. המשמעות היא שאם משרה אחת דרשה 5 כישורים, היא פוצלה ל-5 שורות נפרדות, כאשר נתוני השכר והניסיון שוכפלו לכל שורה. פעולה זו הכרחית כדי לחשב סטטיסטיקות עבור כל כישור בנפרד.
3. טיוב וקיצור שמות: בוצע נרמול של שמות המיומנויות כדי לשפר את קריאות הגרף. שמות ארוכים הוחלפו בקיצורים מקובלים בתעשייה (לדוגמה: "Natural Language Processing" הומר ל-"NLP", ו-"Machine Learning" ל-"ML").
4. אגרגציה וחישוב מדדים: הנתונים קובצו לפי סוג המיומנות, וחושבו שני מדדים מרכזיים לכל כישור:
 - Average_Salary: ממוצע השכר למיומנות.
 - Job_Count: ספירת מספר המופעים (ביקוש) של המיומנות במאגר.
5. סינון רעשים סטטיסטיים: בוצע סינון להסרת מיומנויות נדירות המופיעות פחות מ-10 פעמים במאגר הנתונים. סינון זה נועד להבטיח מובהקות סטטיסטית ולמנוע הטיות הנובעות ממדגם קטן מדי ("outliers").
6. הנדסת מאפיינים לתצוגה: נוצר מדד מחושב חדש בשם Score, המשלב את השכר והביקוש המנורמלים. מדד זה שימש לקביעת סדרי העדיפויות להצגת תוויות הטקסט בגרף, על מנת למנוע עומס ויזואלי ולהציג רק את המיומנויות המשמעותיות ביותר.

ב. שיפור המימוש הויזואלי:

במהלך תהליך הפיתוח והמימוש, בוצעו מספר התאמות לעיצוב המקורי. שינויים אלו נבעו מתובנות שעלו תוך כדי עבודה עם הנתונים האמיתיים, ומתוך צורך לשפר את קריאות הגרף ואת חווית המשתמש.
להלן השינויים המרכזיים שבוצעו:

1. שיפור הקוד הוויזואלי:

- סקאלת צבעים דיפרנציאלית: במקום צבע אחיד או אקראי, הוגדרה סקאלת צבעים רציפה המייצגת את גובה השכר. הסקאלה מדגישה בבירור את הקצוות (שכר גבוה מאוד מול נמוך מאוד) ומציגה צבע ניטרלי עבור ערכי הממוצע, מה שמאפשר זיהוי מגמות במבט מהיר.
 - הוספת מסגרות: נוספו קווי מתאר שחורים ודקים סביב כל נקודה. שינוי זה בוצע כדי למנוע מצב שבו נקודות בעלות צבע בהיר או ניטרלי "נבלעות" ברקע הלבן, ובכך הובטחה ניגודיות ברורה לכל הנתונים.
2. הוספת קווי ייחוס: בניגוד לתכנון הראשוני שעשוי היה להציג רק ענן נקודות, הוספו למימוש הסופי קווים אנכיים ואופקיים המייצגים את החציונים של השכר והביקוש. קווים אלו מחלקים את הגרף לארבעה רביעים ברורים, מה שמעניק למשתמש הקשר מיידי ומאפשר לו להבין האם מיומנות מסוימת נמצאת "מעל" או "מתחת" לשוק הממוצע.
3. מניעת עומס ויזואלי בתוויות: בשל צפיפות הנתונים, זוהתה בעיה של חפיפת טקסטים שהקשתה על קריאת שמות המיומנויות. במימוש הסופי פותחה לוגיקה דינמית למיקום התוויות, המזיזה את הטקסט ביחס לנקודה בהתאם למיקומה במרחב ולצפיפות הסביבתית, כך שהשמות נשארים קריאים וברורים.
4. אופטימיזציה של חווית הבחירה: בממשק הסינון שונתה ברירת המחדל כך שכל רמות הניסיון מוצגות ונבחרות אוטומטית עם טעינת הדף. גישה זו נבחרה כדי להעניק למשתמש תמונת מצב מלאה ורחבה כבר בשלב הראשון, תוך שמירה על הגמישות להסיר או לבחור מחדש רמות ספציפיות בקלות.

ג. אופן המימוש:

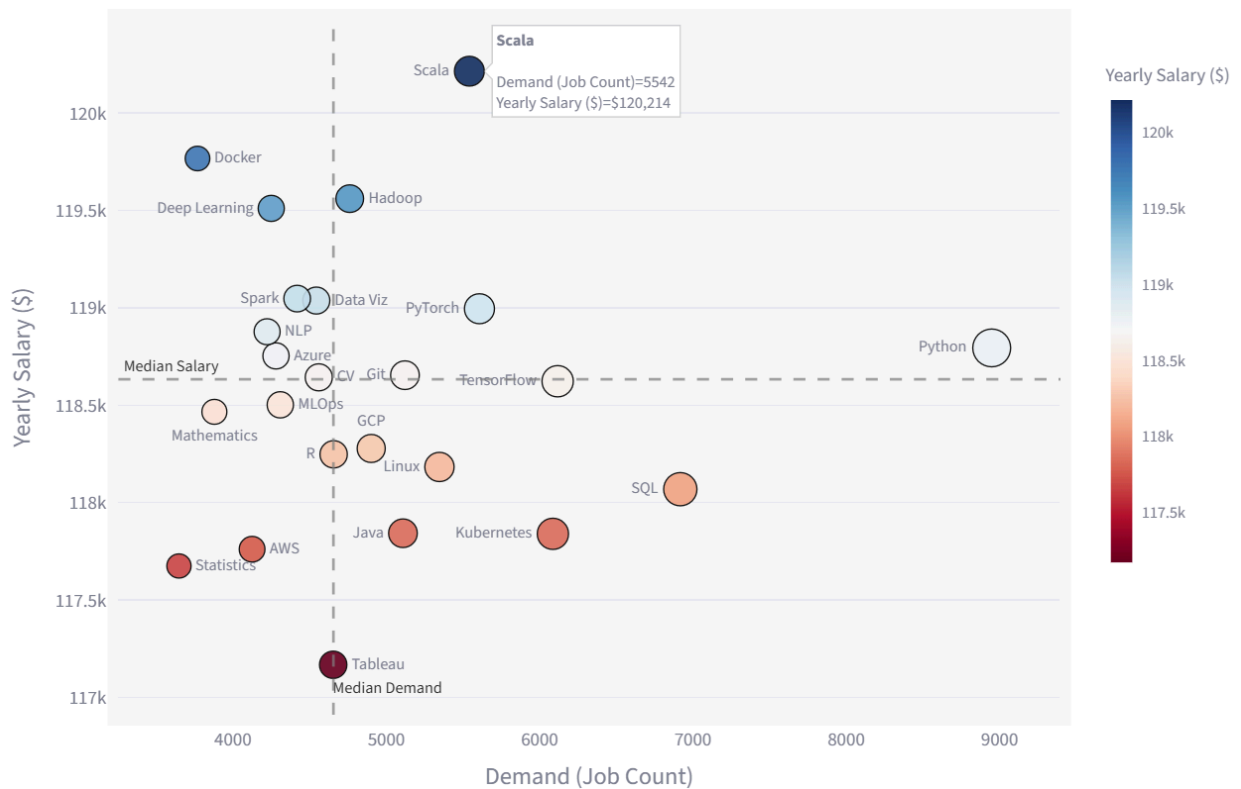
הסבר כללי על הוויזואליזציה:

להלן פירוט המרכיבים הוויזואליים והלוגיקה שמאחוריהם:

1. מערכת הצירים:

- ציר ה-X (ביקוש): מייצג את ה-Job Count, כלומר מספר המשרות במאגר הדורשות את המיומנות הספציפית. ציר זה מעיד על הפופולריות והדרישה למיומנות בשוק.
 - ציר ה-Y (שכר): מייצג את ה-Average Salary, כלומר השכר השנתי הממוצע (בדולרים) המוצע למשרות הדורשות מיומנות זו.
2. קידוד ויזואלי רב-ממדי: כדי להעשיר את המידע המוצג בכל נקודה בגרף, נעשה שימוש בשני ערוצים ויזואליים נוספים מעבר למיקום בצירים:
- גודל הנקודה: גודל הבועה מושפע באופן ישיר מהביקוש (Job Count). בועות גדולות יותר מייצגות מיומנויות פופולריות יותר. בחירה זו נועדה למשוך את עין המשתמש באופן מיידי למיומנויות הדומיננטיות בשוק.
 - צבע: נעשה שימוש בסקאלת צבעים רציפה המייצגת את גובה השכר. קידוד זה מספק אינדיקציה כפולה לשכר (גם המיקום בציר ה-Y וגם הצבע), מה שמקל על זיהוי חריגים (למשל, מיומנות עם ביקוש נמוך אך צבע המעיד על שכר גבוה מאוד).
3. קווי ייחוס וחלוקה לרביעים: על גבי הגרף שורטטו שני קווים מקווקווים המייצגים את החציון של השכר והחציון של הביקוש. הצטלבות קווים אלו מחלקת את המפה לארבעה רביעים אסטרטגיים:
- רביע ימני עליון: מיומנויות "הכוכבות" - גם שכר גבוה מהממוצע וגם ביקוש גבוה.

- רביע שמאלי עליון: מיומנויות נישה יוקרתיות - שכר גבוה אך ביקוש נמוך יחסית.
 - רביע ימני תחתון: מיומנויות בסיס - ביקוש גבוה אך שכר נמוך מהממוצע.
 - רביע שמאלי תחתון: מיומנויות בשוליים - שכר וביקוש נמוכים מהממוצע.
4. אינטראקטיביות ומידע נוסף: הגרף כולל יכולות אינטראקטיביות. במעבר עכבר על כל נקודה, מוצג חלון מידע מפורט הכולל את שם המיומנות המדויק, מספר המשרות המדויק והשכר הממוצע. תכונה זו מאפשרת לשמור על עיצוב נקי מבלי לאבד את הגישה לנתונים הגולמיים המדויקים.

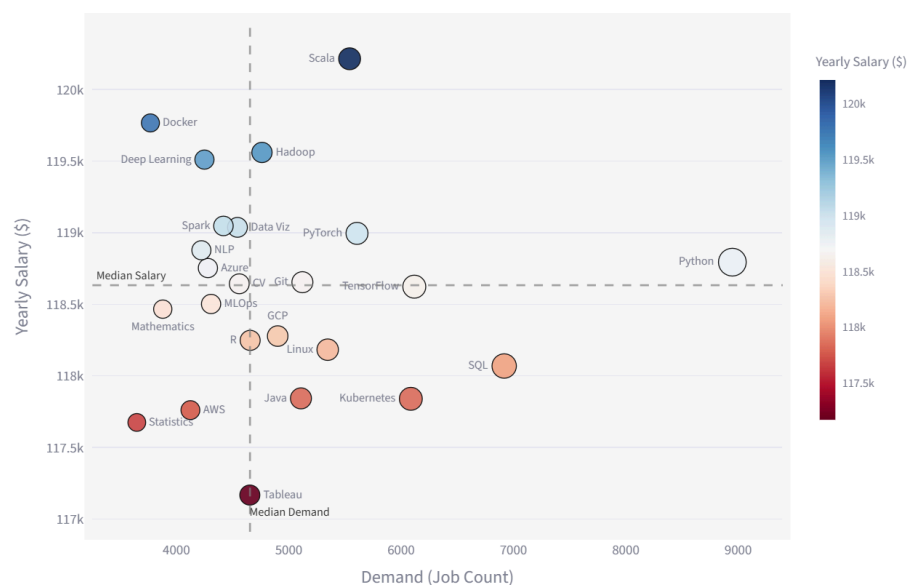


הוויזואליזציה נפתחת עם ברירת מחדל של הצגת כל רמות הניסיון.

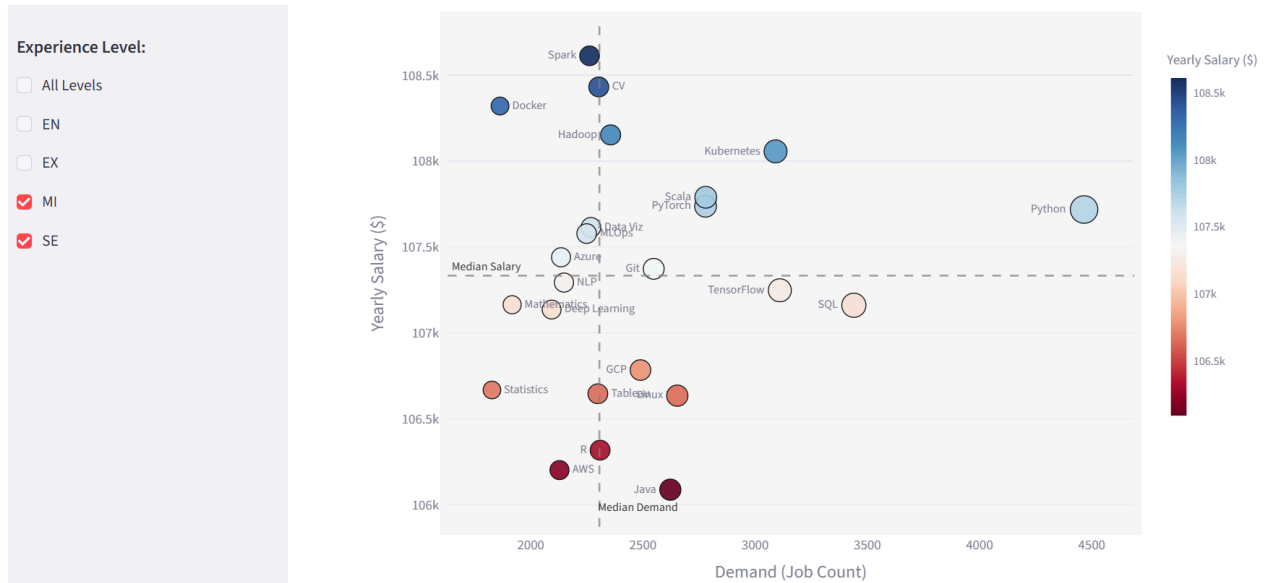
Experience Level:

- ☒ All Levels
- ☒ EN
- ☒ EX
- ☒ MI
- ☒ SE

Showing all experience levels (All Levels is checked).



ניתן לבחור רמות ניסיון ספציפיות שמעניינות אותנו או שילוב של מספר רמות ניסיון.



בנוסף, עבור כל קומבינציה של רמות ניסיון מוצגת טבלה המסכמת את TOP 5 המיומנויות המבוקשות ואת TOP 5 התפקידים בעלי השכר הגבוה ביותר.

Top Salaries

skills_list	Average_Salary
Scala	\$132,720
Deep Learning	\$131,518
Docker	\$130,954
NLP	\$130,869
Hadoop	\$130,730

Highest Demand

skills_list	Job_Count
Python	4483
SQL	3478
TensorFlow	3008
Kubernetes	2996
PyTorch	2829

ד. כתיבת הקוד: בקוד שנכתב עבור וויזואליזציה זו נעשה שימוש בספריות הבאות:

- Streamlit: עבור בניית ממשק המשתמש של אפליקציית Web והצגת בחירות המשתמש להצגת המידע המבוקש לפי רמות הניסיון.
- Pandas: לטובת עיבוד וניהול הנתונים החל מטעינת קובץ הדאטה, דרך סינון לפי שנות ניסיון, ועד לביצוע חישובים מורכבים כמו פיצול רשימות הכישורים וחישוב ממוצעי שכר ומשרות לפי קבוצות.
- Plotly Express: עבור יצירת הגרף האינטראקטיבי שממחיש ויזואלית את הקשר בין השכר לביקוש, ומאפשרת למשתמש לרחף מעל הנקודות כדי לראות מידע נוסף, לסנן לרמות הניסיון שמעניינות אותו ולהתמקד בנתונים.

מטלה 4: האם קיימים הבדלים משמעותיים בשכר בתחום ה-AI בהתאם לתפקיד, אזור גיאוגרפי ורמת ניסיון?

ויוזאליזציה 1: התפלגות שכר לפי תפקיד ורמת ניסיון

א. ביצוע Pre-processing:

ביוזאליזציה זו לא בוצע עיבוד מוקדם מהותי של הנתונים מעבר לניקוי בסיסי הנדרש לצורך הצגה תקינה. הנתונים נטענו מקובץ ה-CSV המקורי כפי שסופק, ובוצעו רק פעולות מינימליות של המרת שדה השכר (salary_usd) לערכים מספריים והסרת רשומות חסרות עבור שדות חיוניים (שם התפקיד והשכר). לא בוצעו אגרגציות, טרנספורמציות, נרמול או סינון מוקדם של הנתונים, על מנת לשמור על נאמנות מלאה להתפלגות המקורית של השכר בשוק העבודה.

ב. שיפור המימוש היוזאלי:

במהלך תהליך המימוש בוצע שינוי ביחס לעיצוב הראשוני של היוזאליזציה. במקור, השוואת השכר בין תפקידי AI התבססה אך ורק על התפלגות השכר לפי תפקיד. עם ההתקדמות בעבודה ועל בסיס תובנות שעלו מהנתונים, הוחלט להוסיף ממד נוסף של רמת ניסיון (experience_level), המוצג באמצעות קידוד צבע בתרשים. שינוי זה מאפשר הבחנה ברורה בין רמות ניסיון שונות (Entry, Mid, Senior, Executive) ומדגיש את השפעת הוותק על פיזור ורמות השכר בכל תפקיד.

בנוסף, נוספה אינטראקציה של סינון לפי סוג העסקה (employment_type), המאפשרת למשתמש לבחון הבדלים בין משרות Full-time, Part-time, Contract וכדומה. הרחבות אלו הפכו את היוזאליזציה מאמצעי תיאורי בסיסי לכלי אנליטי אינטראקטיבי המאפשר חקירה מעמיקה יותר של הנתונים.

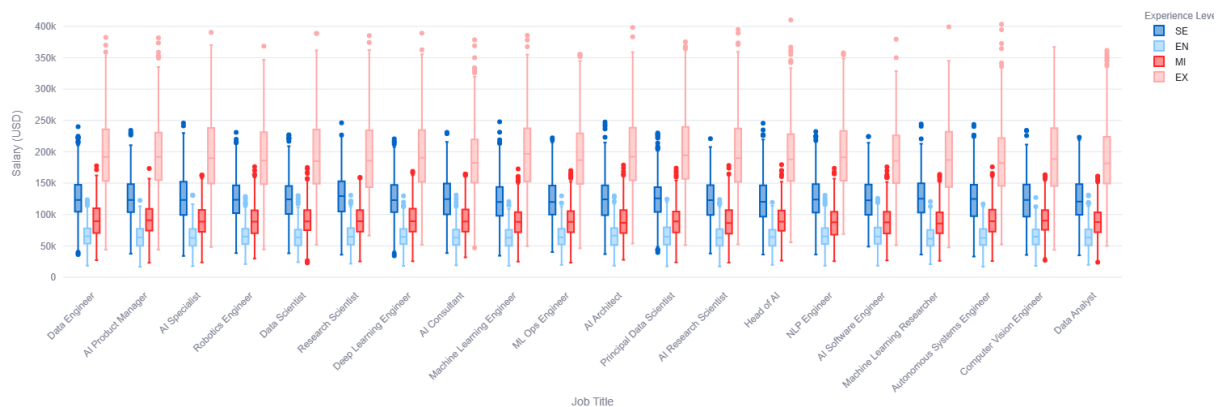
ג. אופן המימוש:

היוזאליזציה מאפשרת למשתמש אינטראקציה באמצעות סרגל צד (Sidebar), שבו ניתן לסנן את הנתונים לפי:

- רמת ניסיון (experience_level)
- סוג העסקה (employment_type)
- מיקום החברה (company_location)

בעת בחירת ערכים במסננים, התרשים מתעדכן בזמן אמת ומציג את התפלגות השכר רק עבור הרשומות הרלוונטיות. בנוסף, מוצגים מדדי KPI (שכר ממוצע, חציון ושכר מקסימלי) המחושבים בהתאם לנתונים המסוננים, ומאפשרים הבנה מהירה של השפעת הבחירות שביצע המשתמש.

Salary Distribution (Box Plot)



Filter Data

Select experience_level

Choose options

Select employment_type

Choose options

ד. כתיבת הקוד:

הוויזואליזציה מומשה באמצעות שפת Python תוך שימוש בספריות הבאות:

- Streamlit – לבניית ממשק אינטראקטיבי, ניהול סרגלי סינון והצגת התרשימים.
- Pandas – לטעינת הנתונים, ניקוי בסיסי וסינון דינמי בהתאם לבחירות המשתמש.
- Plotly Express – ליצירת תרשימי Box Plot אינטראקטיביים, כולל קידוד צבע לפי רמת ניסיון והצגת חריגים.

מבנה הקוד מבוסס על דוגמאות כלליות מתיעוד Streamlit ו-Plotly, אך בוצעו התאמות מהותיות, ביניהן:

- הוספת לוגיקת סינון רב-ממדית
- מיון ציר ה-X לפי חציון שכר
- שילוב מדדי KPI
- התאמת התרשים לצרכים אנליטיים של חקר שוק העבודה בתחום ה-AI

ויזואליזציה 2: Drill-Down לפי מיקום ורמת ניסיון

א. ביצוע Pre-processing:

בויזואליזציה זו לא בוצע עיבוד מוקדם מורכב של הנתונים, אלא עיבוד נקודתי לצורך מימוש ה-Drill-Down. לאחר החלת המסננים הכלליים (כגון רמת ניסיון וסוג העסקה), בוצע סינון נוסף לפי תפקיד נבחר, ולאחריו חישוב שכר חציוני (Median Salary) לכל צירוף של מדינה ורמת ניסיון. בחירה בשכר חציוני נועדה לצמצם את השפעתם של ערכים קיצוניים ולאפשר השוואה יציבה בין מדינות. מעבר לכך, לא בוצעו טרנספורמציות או שינויי מבנה בנתונים.

ב. שיפור המימוש הויזואלי:

העיצוב המקורי התמקד בהשוואת שכר גלובלית בין תפקידים, אך במהלך המימוש עלה הצורך לאפשר חקירה ממוקדת יותר של השפעת המיקום הגאוגרפי על השכר עבור תפקיד מסוים. לכן, הויזואליזציה הורחבה למבנה של Drill-Down, שבו המשתמש בוחר תפקיד ספציפי, והמערכת מציגה עבורו את השכר החציוני במדינות שונות, תוך הבחנה בין רמות ניסיון.

בנוסף, הוחלט להוסיף סרגל שליטה המאפשר קביעה דינמית של מספר המדינות המוצגות, על מנת למנוע עומס ויזואלי ולאפשר מיקוד במדינות הרלוונטיות ביותר. שינוי זה שיפר משמעותית את הקריאות ואת היכולת לבצע השוואות בין מדינות בעלות רמות שכר שונות.

כמו כן, נוספה אפשרות למיון המדינות בציר האנכי לפי רמת ניסיון נבחרת. המיון מתבצע על-פי השכר החציוני של אותה רמת ניסיון בכל מדינה, כך שהמדינות מסודרות מהשכר הגבוה לנמוך. כברירת מחדל, המיון מתבצע לפי רמת EX (Executive), המייצגת את הדרג הבכיר ביותר, אך המשתמש יכול לבחור למיין גם לפי רמות ניסיון אחרות (Senior, Mid או Entry) לצורך השוואה וניתוח מעמיק יותר של דפוסי השכר בין מדינות.

ג. אופן המימוש:

הויזואליזציה תומכת במספר שכבות אינטראקציה המאפשרות חקירה דינמית של נתוני השכר: בחירת תפקיד באמצעות תיבת בחירה (Select Box). קביעת מספר המדינות המוצגות באמצעות סרגל הזזה (Slider). הצגת השכר החציוני לכל מדינה, כאשר הנקודות מקודדות בצבע לפי רמת ניסיון. בנוסף, נוספה אפשרות למיין את המדינות בציר האנכי לפי השכר החציוני של רמת ניסיון נבחרת, כאשר כברירת מחדל המיון מתבצע לפי רמת EX. בעת שינוי אחד מהפרמטרים, התרשים מתעדכן בזמן אמת ומאפשר למשתמש לבחון כיצד מיקום גיאוגרפי ורמת ניסיון משפיעים יחד על השכר בתפקיד הנבחר.

Median Salary by Country and Experience Level

Select job title

AI Architect

Sort countries by salary of experience level

EX



ד. כתיבת הקוד:

המימוש בוצע באמצעות Python תוך שימוש בספריות:

- Streamlit – לניהול אינטראקציות משתמש (בחירת תפקיד, סרגל כמות מדינות, פריסת עמודות).
- Pandas – לסינון הנתונים, קיבוץ לפי מדינה ורמת ניסיון וחישוב שכר חציוני.
- Plotly Express – ליצירת תרשים פיזור (Scatter Plot) אינטראקטיבי, כולל קידוד צבע לפי רמת ניסיון.

הקוד מבוסס על עקרונות ודוגמאות כלליות מתיעוד Plotly ו-Streamlit, אך הותאם לצורכי העבודה באמצעות:

- קביעת סדר קטגוריאלני לרמות ניסיון
- הוספת קווי Dumbbell המחברים בין ערכי מינימום ומקסימום בכל מדינה
- שליטה דינמית בכמות המדינות המוצגות
- שילוב הוויזואליזציה כחלק מתהליך Drill-Down רב-שלבי

ספריות חיצוניות ושימוש בAl:

עבור העבודה נעשה שימוש בספריות החיצוניות:

- Streamlit - לניהול אינטראקציות משתמש
- Pandas - לסינון הנתונים
- Plotly Express - ליצירת התרשימים

כל וויזואליזציה עשתה שימוש בכלים אלו באופן המותאם לה.

נעשה שימוש בAl באופן הבא: תחילה חשבנו מה נרצה שהוויזואליזציה תציג במדויק וכיצד תראה. לאחר מכן שלחנו לAl את הדרישות המדויקות ליצירת הוויזואליזציה - לרבות הכלי בו נעבוד, ממה מורכבת הוויזואליזציה במדויק, סקאלת צבעים, האופן האיטרטיבי בו יעבוד וכו' - כלומר תיאור מקיף של כל מה שיהיה בוויזואליזציה כך שהAl יממש את הקוד שמבצע את כל הדרישות של הוויזואליזציה.