

# השפעת ה-AI על שוק העבודה

## ייזואלייזציה של מידע

דו"ח מסכם

### מגישיים:

נעה שבב - 210025631

חן מרכוביץ - 211389374

סתוי עזרי - 206360067

ניב דולב - 323817866

### תאריך הגשה:

23.01.2026

### קישור לאפליקציית הדashboard:

<https://data-visualization---projectapp-xdv3vriyf7wojgvtqqlwva.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 ביזואליזציה של נתונים אודוט השפעת AI על שוק העבודה העולמי. בשנים האחרונות הפך ה-AI לאחד הגורמים המשמעותיים ביותר המשפיעים על שוק העבודה העולמי. שילובו של AI בתהליכי עבודה וקבלת החלטות, שינה באופן רחב את מבנה השירות, ואת הציפיות מהעובדים. מטרת העבודה היא לבחון בצורה ייזואלית וمبוססת נתונים את השפעת AI על שוק העבודה, וזה השאלה המרכזית אליה אנחנו רוצים לענות.

לטובת עבודה זו נעשה שימוש בדאטא שנלקחה מאתר Kaggle : Global AI Job Market & Salary Trends 2025

- בתחום ה-AI מדיניות רבות בעולם. הדאטא כולל פרטיים כגון: תפקיד, שכר, רמת ניסיון, סוג העסקה, גודל החברה, רמת עבודה מרוחק, מקום גיאוגרפי ועוד.
- נרצה לענות על מספר שאלות משנה נוספתו בפרויקט:
1. האם קיימים הבדלים גיאוגרפיים מובהקים באימוץ AI ברחבי העולם?
  2. ניתוח השוואתי רב-מדדי של תפקידים עבודה בתחום ה-AI.
  3. השוואה בין רמת הביקוש למיניות 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

הdatapset שבחרנו לעבוד איתה מכילה 30,000 רשומות ו-19 פיצרים, כאשר כל רשומה מייצגת מודעת עבודה אחת בתחום ה-AI. לכל פריט יש סט של מאפיינים המתארים את התפקיד, השכר, דרישות הניסיון, המיקום והתנאים הנלוויים.

מיofi הנתונים בהתאם לטיפוליה האבסטרקטית של Munzner :

dataset Type : הנתונים הינם בטבעיות עם מפתח אחד לכל item ולכל הסוג הוא FLAT TABLE : Data types

- כל item בנתונים שלו מייצג מודעת דרישים של משרה בתחום ה-AI.
- בטבלה מגוון שדות קטגוריאליים, אודינליים, 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	הمدينة בה המשרה מוצעת	20
company_size	Ordinal	גודל החברה	3
employee_residence	Categorical	מקום מגורי העובד	50
remote_ratio	Quantitative	רמת התאמת המשרד לעבודה מרוחק	3
required_skills	Categorical	כמעט כל רשומה עם רשימת מיומנויות "יחודית"	25,702
education_required	Ordinal	High School / Bachelor / Master / PhD	4
years_experience	Quantitative	טוווח שנות ניסיון שונים (0–20)	20
industry	Categorical	ענפים שונים	15
posting_date	Ordinal	תאריכי פרסום שונים	486
application_deadline	Quantitative	DDL'ין שונים	543
job_description_length	Quantitative	ערכים שונים (אורך טקסט המשרה)	2,000
benefits_score	Quantitative	ערכי ציון היטבות 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, סוג השימוש והשפעה על עובדים ופרודוקטיביות.

המשימה משלבת בין נתוני הדטה לבין מידע גאוגרפי, ומטרתה לזהות דפוסים מרחביים באימוץ AI.  
במונחים של munzner:

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

### **מטלה 2: ניתוח השוואתי רב-מדדי של תפקידים של תפקידי עבודה בתחום ה-AI**

**במונחים של התחום:** מטלה זו עוסקת בניתוח השוואתי רב-מדדי של תפקידים של תפקידי עבודה בתחום ה-AI, במטרה לאמוד את ה"אטראקטיביות" של שירותי מנכודות מבט מגוונות ולא על סמך מידע בלבד. באמצעות ייזואלייזציה אינטראקטיבית, המערכת מאפשרת למשתמשים לחקור קשיים ודפוסים בין תפקידים (כגון Data Scientist או ML Engineer) על בסיס חמשה מאפייני מפתח: שכר, ניסיון נדרש, היקף ביצוע, הטבות ופיזור תעשייתי. הניתוח נעוד לחשוף פערים וחירגים בשוק, למשל, תפקידים בעלי שכר גבוה אך דרישות סף גבוהה או היעץ שירות נמוך ובכך לספק למשתמש תמונה הוליסטית ומורכבת המסייעת בקבלת החלטות קריירה מושכלות.

במונחים של munzner:

Action	Sub-action	Target	Explanation
analyze → consume	present	distribution	הציג הפיזור של תפקיד AI שונים על פני מספר מדדים מרכזיים (שכר, ביצוע, 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 ביחס נמוך, או תפקידים עם ביחס רחוב אך תגמול ושורות הטבות מצומצמים.
query	compare	trends		השוואה תפקידי AI שונים תחת מדרדים שונים, למשל השוואת בין תפקידי מחקר לתפקיד פיתוח מבחינת שכר ניסיון נדרש, או בין תפקיד מוצר לתפקיד תשתיות מבחינת ביחס ופייזור תעשייתי.
query	summarize	patterns		זיהוי דפוסים כלליים בשוק העבודה, כגון תפקידים המאפיינים באטרקטיביות גבוהה במספר מדרדים במקביל לעומת תפקידים המצטננים במדד אחד בלבד אך חלשים אחרים.

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

**במנוחים של התחום:** המשימה בוחנת את הקשר בין ביקוש למיומנויות טכנולוגיות שונות בתחום ה-AI (כפי שמשמעותה בנסיבות השירותים כל מיומנות) לבין רמת השכר השנתית הממוצעת. ניתוח זה מאפשר לזהות מיומנויות מבוקשות במיוחד, מיומנויות מתgelמות כלכלית, והאם קיימן פער בין ביקוש גבוה לתגמול כספי. מטרת המשימה לחושוף דפוסים ומינימיות חריגות בשוק העבודה.

**במנוחים של munzner:**

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

query	compare	trends	השוואת מגמות בין קבוצות מiomנויות שונות, כגון שפות תונות, כל תשתית וטכנולוגיות למדת מכונה, כדי לבחון כיצד הביקוש משפיע על רמת השכר.
query	summarize	patterns	זהוי דפוסים כלליים, למשל מiomנויות הממוקמות מעל החזון ה- <i>n</i> בבדיקה והן בשכר, לעומת מiomנויות הנמצאות בערך בין ביקוש גבוה לתגמול כספי נמוך.

#### טבלה 4: האם קיימים הבדלים משמעותיים בשכר בהתאם לתפקיד, אזור גיאוגרפי ורמת ניסיון?

במונחים של התחום: בטבלה זו נבחנן כיצד מאפיינים מסוימים ותעסוקותיהם משפיעים על רמת השכר בשוק העובדה בתחום ה-*n*.

במציאות ניתן שכר, נבחנן 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	הציג הקשר בין שכר לבין תפקיד, אזור גיאוגרפי ורמת ניסיון באמצעות ויזואלייזציות השוואתיות
search	locate	outliers	זהוי משרות או אזורים עם שכר חריג במיוחד (גובה או נמוך)
query	compare	trends	השוואת מגמות שכר בין תפקידים שונים, אזורים שונים ורמת ניסיון שונות
query	summarize	patterns	זהוי דפוסים כלליים בשוק העבודה של <i>n</i> , כגון שילובי תפקיד אזור ניסיון המניבים שכר גבוה במיוחד

### חלק שלישי: עיצובים חלופיים

מציג את העיצובים החלופיים עבור כל מטרה שהציגנו:

**מטרת 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, סוג השימוש והשפעה על עובדים ופרודוקטיביות.

המשימה משלבת בין נתוני הדатаה לבין מידע גאוגרפי, ומטרתה לזרות דפוסים מרחבים באימוץ AI.

**במונחים של munzner:**

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

ב. סקיצה:

**חלופה 1:**

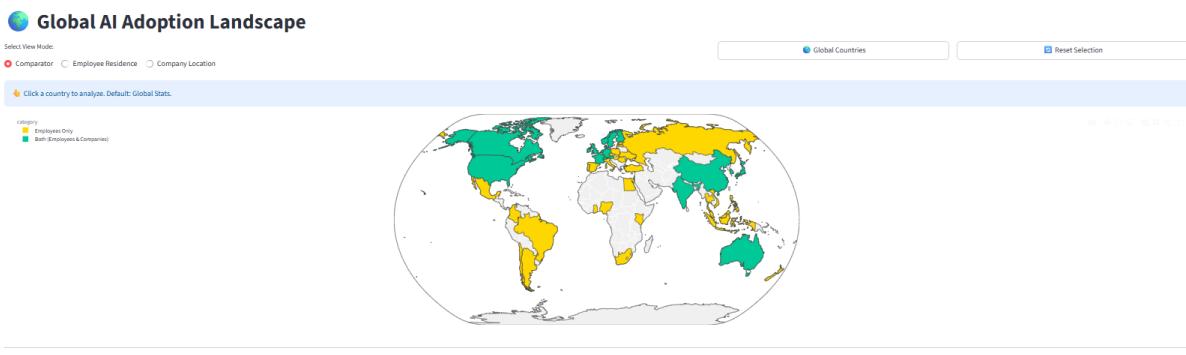
תיאור החלופה: הדשبورד מורכב מפה עולם אינטראקטיבית (Choropleth Map).

- צבע המדינה משתנה בהתאם לכמות המשרות/העובדים (סקאלת צבעים רציפה).

- לחיצה על מדינה במפה מסננת את הנתונים ומציגה בצד הדשبورד 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 הchlופה במונחים של Marks & Channels

- **Marks:**
  - Areas (surfaces): גבולות גיאוגרפיים של מדינות (2D Area).
  - Color Saturation (רויית צבע): משתנה כמותי (מספר משרחות/עובדים). כהה יותר = יותר משרחות.
  - Spatial Region (מקום מרחב): משתנה קטגוריאלי (זהות המדינה).
  - Interaction (אינטראקטיה): Select/Click לbijoux סינון (Filter).

## ד. אקספרסיביות ואפקטיביות

- אקספרסיביות: הchlופה מציגה את כל המדינות הקיימות בDATA. השימוש במפה הוא ייצוג טבעי למידע גיאוגרפי. מצב ההשוואה (Comparator) חושף מידע השוואתי שלא ניתן לראותם במבט יחיד על מפה.
- אפקטיביות:
  - ערך המיקום המרחבי הוא האפקטיבי ביותר לזהות מדינות.
  - ערך הצבע (Color Saturation) פחות מדויק להשוואה כמותית עדינה (קשה להבחין בהבדל בין גוון 60%-70%), אך מציין לזהות מגמות מקרים (איפה יש "רבה" ואיפה "מעט").

## ה. יתרונות וחסרונות

- יתרונות:
  1. אינטואיטיביות: מפה היא הדרך המהירה ביותר למשתמש להבין הקשר גיאוגרפי ללא צורך בלימוד המערכת.
  2. צלילה לפרטים (Drill-down): המפה מספקת תמונה על, והאינטראקטיה מאפשרת צלילה לפרטים (Drill-down) מוביל להעמקה על המסר הראשוני.
  3. הקשר מרחביבי: מאפשרת לזהות דפוסים אזוריים (למשל, גוש מדינות באירופה עם אימוץ גבוה) שלא היו נראים בגרפ עמודות רגיל.

## חסרונות:

1. **Area Bias** (הטיית שטח): מדינות גדולות פיזית (כמו רוסיה או קנדה) תופסות הרבה תשומת לב

ויזואלית (די) גם אם הנתונים שלן נמוכים, בעוד מדינות קטנות עם נתונים גבוהים (כמו ישראל או סינגפור) עלולות להיעלם.

2. **דיק נמור בהשוואה:** קשה להשוות ערכיהם כמותיים מדויקים באמצעות גווני צבע בלבד (אם צרפת כהה יותר מגרמניה?).

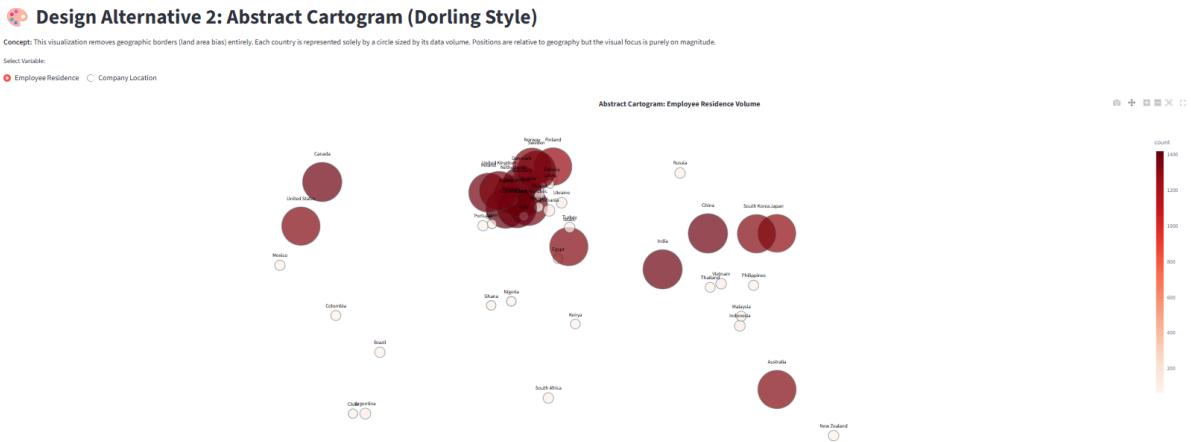
## חולופה 2: קרטוגרפיה מופשטת (Dorling Cartogram) - העיצוב שלא נבחר

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

### a. תיאור המטלה (Task Abstraction)

Action	Sub action	Target	Explanation
analyze → consume	present	distribution	מצגת התפלגות גלובלית של משרות ועובדים על גבי מפה גיאוגרפית.
search	locate	outliers	זהוי מדינות בעלות פעילות חריפה (גובהה מאוד או נמוכה) ביחס לשאר העולם.
query	compare	trends	השוואה בין מדינות שונות (Comparator Mode) לבחינת הבדלים בשכר ובסוגי התפקדים.
query	summarize	patterns	קבלת תמונה מצב כללי על ריכוז התעשייה (למשל: צפון אמריקה מול אירופה).

## ב. סקיצה



## ג. ניתוח Marks & Channels

1. **Mark (סימן): נקודה/שטח (Point/Area).** העיגול הוא היחידה הגרפי המיצגת מדינה.
2. **Channel 1 (ערוך גודל): שטח דו-ממדי (2D Area).** שטח העיגול מקודד את המשתנה הכמותי (מספר השירותים). זהו הערך המרכזי והדומיננטי ביותר בחלוקת זו.
3. **Channel 2 (ערוך מיקום): מיקום מרחבי (Spatial Position).** המיקום הוא יחס (טופולוגי) ולא אבסולוטי. הוא נדרש לשמר את ה"מודל המנטלי" של מפת העולם (אurope במרכז, ארה"ב משמאל) אך אינו מודיעין גיאוגרפיה.
4. **Channel 3 (צבע - אופציוני):** ניתן להשתמש בצבע לקידוד מידע נוסף (כגון שכר ממוצע) או כ-Redundant Coding לחיזוק הקריאות.

## ד. עקרונות האקספריביות והאפקטיביות

- **אקספריביות:** החלופה אקספרטיבית מאוד לנוטוי Magnitude. היא מבטלת את "פקטור השקר" (Lie Factor) של מפות כורופלת (Choropleth), שבהן שטח גיאוגרפי ריק תופס נפח ויזואלי ומעוות את תפיסת המידע. בחלוקת זו, רק מידע רלוונטי (Data) תופס מקום במסך (Ink).
- **אפקטיביות:**
  - יתרון: עיליה מאוד לזייהו "מי השחקנים המרכזיים" ללא רושי רקע.
  - חיסרון: על פי Stevens' Psychophysical Power Law, בני אדם מתकשים להעריך יחס שטחים בצורה ליניארית (אם נתונים להערך שטח בחסר לעומת אורך). לכן, המשתמש יתקשה לדעת אם עיגול א' גדול בבדיקה פ' 2 מעיגול ב'. כמו כן, האפקטיביות של מטלת Locate (איתור מדינה ספציפית) נפגעת ממשמעותית בהעדר גבולות מוכרים.

ה. יתרונות וחסרונות

• יתרונות:

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

חסרונות:

1. **אובדן Kontext (Loss of Context):** המשתמש מאבד את הקשר הגיאוגרפי המדוייק. קשה להזיהות מדינות שאין בולטות במיוחד (עיגולים קטנים) ללא קווים מתאר של יבשת שיעזרו בההתמצאות.
2. **עומס קוגניטיבי בזיהוי:** ללא תוויתות ברורות (Labels) על כל עיגול, המשתמש יתנסה להבין איזה מדינה מיצגת על ידי איזה עיגול, למעט המדינות המוכרות ביותר במיקומן המשוער.

הצדקת הבחירה בחלופה הסופית:

- אינטואיטיביות ומוכנות: החלופה מציגה את הנתונים על גבי מודל מנטלי מוכר (מפת העולם), מה שמאפשר למשתמש להתמצא למרחב באופן מיידי ללא צורך בלמידה ייצוג חדש ומוספט.
- איתור מהיר (Locate): ניתן לאתר מדינות ספציפיות בקלות על בסיס צורתן ומקום הגיאוגרפיה, בניגוד לקרטוגרפיה המופשטת שבה המיקום היחסי משתבש והזיהוי הופך למאצ' קוגניטיבי.
- שימור הקשר מרחב: החלופה משמרת את הטופולוגיה (שכנות בין מדינות), מה שמאפשר זיהוי דפוסים ומגמות אזוריות (למשל: ריכוז גבוה באירופה) בצורה טובה יותר מאשר בועות מנוקחות.
- גישה לנ נתונים מדוייקים: אינטראקטיביות hover מאפשרת גישה לערבים המספריים המדוייקים, ובכך מפיצה על החישורן של ערך הצבע (Color Saturation) בהשוואה כמותית עדינה.
- עקרונות Munzner: החלופה ממחשת את עקרון השימוש ב-Spatial Region (מיקום מרחב) כערוך האפקטיבי ביותר לנ נתונים קטגוריאליים (זהות המדינה), תוך העדפת קלות הזיהוי על פני דיוק בהשוואה שטחים (שבו בני אדם מתקשים לפי חוק Stevens).
- מניעת עומס ויזואלי: המפה מספקת "עוגן" ויזואלי ברור, בעוד שאוסף רב של עיגולים מרוחפים (בחילופה השניה) עלול ליצור עומס ויזואלי ובלבול כאשר מדובר במספר רב של מדינות.

הקרטוגרפיה המופשטת (Dorling Cartogram) נשארת חילופה תקפה להציג סדר גודל נקיים ללא "הטיית שטח" (Area Bias), אך לא נבחרה כמיושה הסופי בשל הפגיעה המשמעותית ביכולת ההסתמאות ובזיהוי המהיר של מדינות ללא תוויות טקסט מרובות.

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

מטלה זו עוסקת בניתוח השוואתי רב-מדדי של תפקידי עבודה בתחום ה-AI, במטרה לאמוד את ה"אטראקטיביות" של שירותי שונות מנוקדות מבט מגוונות ולא על סמך מדד אחד. באמצעות ייזואלייזציה אינטראקטיבית, המערכת מאפשרת למשתמשים לחזור קשרים ודפוסים בין תפקידים (כגון Data Scientist או ML Engineer) על בסיס חמישה מאפייני מפתח: שכר, ניסיון נדרש, היקף ביקוש, הטבות ופייזור תעשייתי. הניתוחನועד לחשוף פערים וחיריגים בשוק, למשל, תפקדים בעלי שכר גבוה אך דרישות סף גבוהה או היצע שירותי נמוך ובכך לספק למשתמש תמונה הוליסטית ומורכבת המסייעת בקבלת החלטות קריירה מושכלות.

Action	Sub-action	Target	Explanation
analyze → consume	present	distribution	ցגת הפיזור של תפקידי AI שונים על פני מספר מדדים מרכזיים (שכר, ביקוש, ניסיון נדרש, הטבות ותעשייה), במטרה להבין את המיקום הייחודי של כל תפקיד בשוק העבודה.
search	locate	outliers	זיהוי תפקידי AI חריגים, כגון תפקידים עם שכר גבוה במיוחד ביחס נמוך, או תפקידים עם ביקוש רחוב אך תגמול ושורת הטבות מצומצמים.
query	compare	trends	השוואה תפקידי AI שונים תחת מדדים שונים, למשל השוואת בין תפקידי מחקר לתפקיד פיתוח מבחינת שכר וניסיון נדרש, או בין תפקיד מוצר לתפקיד תשתיות מבחינת ביקוש ופייזור תעשייתי.
query	summarize	patterns	זיהוי דפוסים כלליים בשוק העבודה, כגון תפקידים המאפיינים אטרקטיביות גבוהה במספר מדדים במקביל לעומת תפקידים המצטיינים במסד אחד בלבד אך חלשים אחרים.

### חולופת עיצוב 1: תרשימים מעודות מקובץ (Grouped Bar Chart) עם מפתח חום משלימה

א. המטלה עליה עונה החלופה

במונחים אבסטרקטיביים:

השוואה, זיהוי חריגים וסיכום דפוסים (Compare, Locate, Summarize)

במונחי התחום:

השוואה תפקידי AI על פני מספר מדדים מרכזיים לצורך הערכת האטרקטיביות הייחוסית שלהם בשוק העבודה.

## ב. סקירה

הסקירה מציגה את תפקידיה העיקריים על הציר האופקי, ולצדם קבוצות של עמודות. כל עמודה מייצגת ממד שונה (כגון שכר, היקף משירות או ניסיון נדרש), כאשר הערכים מנורמלים לטווח אחיד לצורר השווואה בין ממדים בעלי יחידות שונות.

מתחת לתרשים מוצגת טבלה משלימה עם קידוד צבעים, המציג את הערכים הגולמיים של אותם ממדים.

### 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

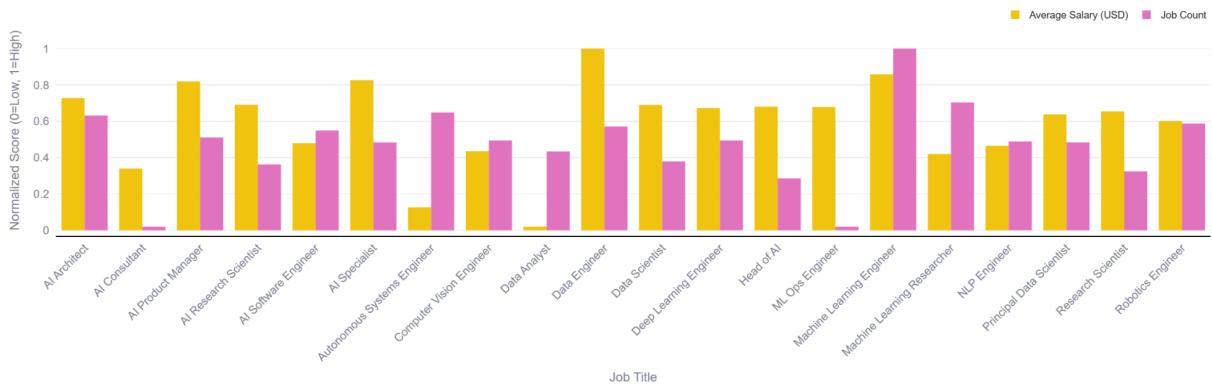
Average Salary (USD)  Job Count  Years of Experience  Benefits Score  Remote Work Ratio (%)

#### 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. תרשימים העמודות:

:Marks

עמודות (מלבינים)

:Channels

- מיקום אופק' - תפקיד עבודה
- גובה העמודה - ערך ממוצע מנורמל
- צבע (קטגוריאלי) - סוג המדד
- קיבוץ - שירט 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 הכלומי של המדד
- טקסט- הערך המספרי המדויק

## ד. אקספרסיביות ואפקטיביות

### אקספרסיביות:

כל עורך ויזואלי מיצג משתנה נתונים רלוונטי באופן ישיר וברור, ללא הצגת מידע שאינו קיים נתונים. בתרשים העמודות, הנירמול 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 תפקידיים. הנירמול תומך בהשוואה רב-מידית, כאשר אינטראקטיביות `hover` מספקת גישה לערכים האבסולוטיים ומפיצה על אובדן קנה המידה המקורי. מפת החום משלימה את התרשים בכך שהיא מאפשרת זיהוי מהיר של ערכים גבוהים ונמוכים ושומרת על דיקט 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 בהשואות ממוחזקיות

חסרונות:

- הנירמול עלול להסתיר הבדלים אבסולוטיים ללא שימוש באינטראקטיבית
- עומס ויזואלי עולה עם ריבוי מדדים
- מפת החום אפשרית לפחות עמודה ולא לפחות הטבלה(סקלה שונה לכל פיצר והערכים ללא נורמל)

## חלוות עיצוב 2: תרשيم קוואורדינטות מקבילות (Parallel Coordinates)

א. המטלה עליה עונה החלופה

- אבסטרקטיות:

השוואה רב-ממדית וודחוי דפוסים וחיריגים  
(Compare, Summarize, Locate outliers)

ב. מונחי התחום:

השוואת תפקידי AI על פני שכר, ניסיון, ביקוש, הabytes, פיזור תעשייתי ועבודה מרוחק.

ב. סקיצה

הסקיצה כוללת מספר צירים אנכיים מקבילים, כאשר כל ציר מייצג מדד ממוחזקי אחד.  
כל תפקיד עבודה מיוצג על ידי קו המחבר בין ערכי על פני הציריהם השונים.



ג. Marks & Channels

:Marks

- קווים (קו אחד לכל תפקיד)
- צירים אנכיים :Channels
- מיקום לאורך הציר - ערך ממוחזקי
- צבע רציף - שכר ממוצע

- סדר הציגים - סוג המدد

ד. אקספרסיביות ואפקטיביות

- אקספרסיביות:

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

- אפקטיביות:

מיקום הוא ערוץ אפקטיבי להשוואת ערכיהם, אך חפיפות בין קווים מקשوت על השוואה מדויקת בין תפקדים בודדים.

בנוסף, זיהוי של תפקיך מסוים דורש אינטראקציה משלימה או טבלה נלוית.

ה. יתרונות וחסרונות

יתרונות

- אפשר הצגה של מספר רב של מדדים בו-זמנית

- מתאים לזיהוי דפוסים כלליים ויחסי trade-off בין מדדים

חסרונות

- קושי בזיהוי חד-משמעות של תפקיך ספציפי

- חפיפות רבות יוצרות עומס ויזואלי

- דיק נמוך בהשוואה נקודתית של ערכים

לאחר בחינת שתי החלופות, נבחרה חלופה תרשימים העמודות כמיושם הסופי של הפרויקט מכיוון:

- מציגה נתונים באופן פשוט ונוח למשתמש

- ניתנת הבין על איזה משרה מדובר בקלות

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

- הבחירה מספקת תמצאה טוביה יותר בנסיבות השוואה בין תפקדים בודדים

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

- בהשוואה לקו-ודינומות מקבילות, מתאפשר ייצוג פחות עמוס וברור יותר למשתמש

תרשימים הקואורדינטות המקבילות נשאר חלופה תקופה לחקירה ראשונית, אך לא נבחר כמיושם הסופי בשל מגבלותיו בזיהוי והשוואה מדויקת של תפקדים.

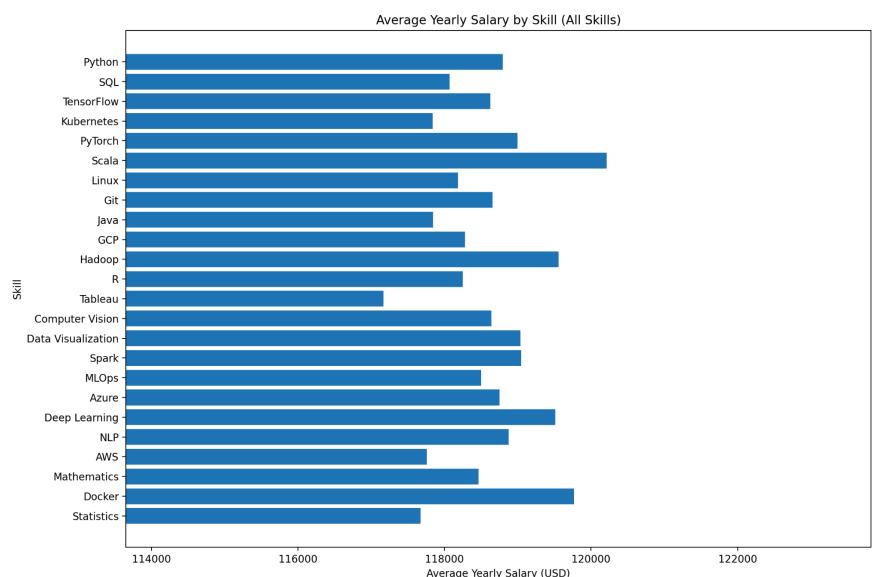
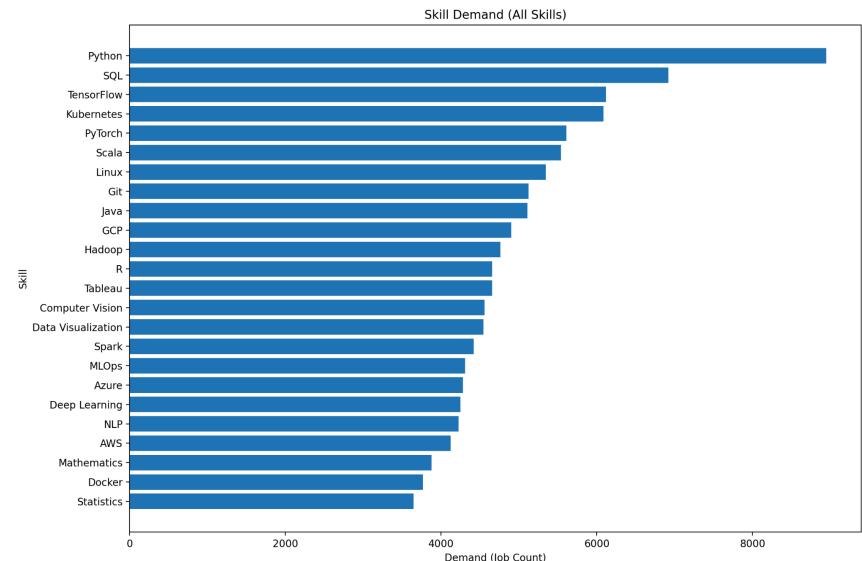
**מטלה 3 : השוואת בין רמת הביקוש למיומנויות שונות בשוק העבודה בתחום ה-AI לבין רמת השכר המוצעת עבורן.**

א. תיאור המטלה במנוחים אבסטרקטיים ובמנוחה התחום עבור שתי הحلופות:  
**במנוחים של התחום:** המשימה בוחנת את הקשר בין ביקוש למיומנויות טכנולוגיות שונות בתחום ה-AI (כפי שמשתקף בקטגוריות הדורשות כל מיומנות) לבין רמת השכר השנתית המוצעת. ניתוח זה מאפשר לזהות מיומנויות מובוקשות במיוחד, מיומנויות מתגמלות כלכלית, והאם קיימים פער בין ביקוש גבוה לתגמול כספי. מטרת המשימה לחשוף דפוסים ומיומנויות חריגות בשוק העבודה.  
**במנוחים של nerzhunchnun:**

Action	Sub action	Target	Explanation
analyze → consume	present	distribution	הציג פיזור המיומנויות על פני ציר הביקוש (מספר משרחות) והשכר השנתי, במטרה להבין את המיקום היחסי של כל מיומנות בשוק העבודה.
search	locate	outliers	זהוי מיומנויות חריגות, כגון מיומנויות עם שכר גבוה אך ביקוש נמוך או מיומנויות עם ביקוש גבוה אך שכר ממוצע נמוך יחסית
query	compare	trends	השוואה מגמות בין קבוצות מיומנויות שונות, כגון שפנות תכנות, כל תשתית טכנולוגיות למידת מכונה, כדי לבחון כיצד הביקוש משפיע על רמת השכר.
query	summarize	patterns	זהוי 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:

- גרפ' ראשון: ציר X - ביקוש (מספר משרות), ציר Y - מיומנויות

- גרפ' שני: ציר X - שכר שנתי ממוצע, ציר Y - מיומנויות

הגרפים מוצגים זה מתחת לזה ומשתמשים באותו סדר מיומנויות לצורכי השוואה ויזואלית. הגרפים הנ"ל

ללא אינטראקטיביות, בעלי ייצוג סטטי.

## ג. תיאור החלוקת במונחים של Marks & Channels:

:Marks

Bars (מלבנים) -

:Channels

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

(Position / Skill) (ציר Y): משתנה קטגוריאלי -

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

## ד. אקספרסיביות ואפקטיביות

אקספרסיביות:

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

- אין הצגה של מידע שלא קיים בדתא

אפקטיביות:

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

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

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

## ה. יתרונות וחסרונות:

יתרונות:

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

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

- ההפרדה לשני גרפים מאפשרת מיקוד בכל משתנה בנפרד (ביקורת או שכר), ומונעת ערבות בין משתנים בעלי יחידות מידת שונות.

- היצוג הקטורי עלי ציר ה-Y תומך בהשוואה ישירה בין מיומנויות רבות, גם כאשר מספר הקטגוריות גבוהה יחסית.

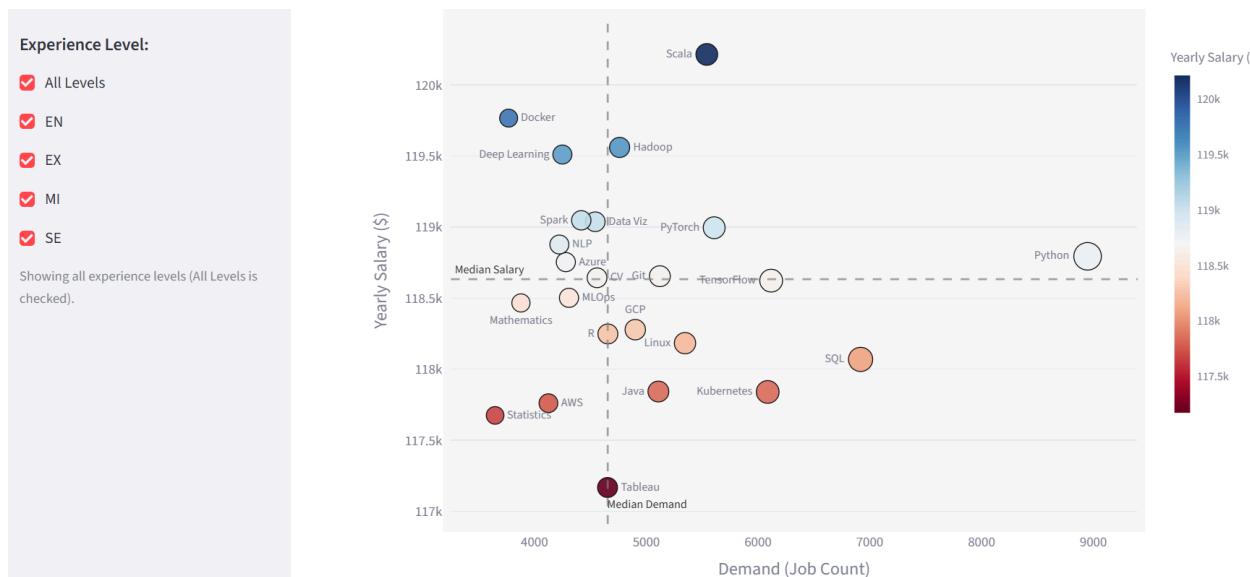
חסרונות:

- לאחר שכל גרף מייצג משתנה אחד בלבד, אין קידוד משותף לשני משתנים, ולכן לא ניתן לזרות קשרים או יחסים בין ביקוש לשכר באמצעות ערכאים וייזואליים, אלא רק באמצעות השוואה קוגניטיבית בין גרפים נפרדים.

- היעדר שימוש בערכים נוספים כגון color או size מגביל את יכולת הוייזואליזציה לעבר מידע רב-ממדדי.

- דוחי חריגים (למשל מיומנויות עם ביקוש נמוך אך שכר גבוה) אינם נתמך ישירות על ידי הערכים הויזואליים, ודורש קריאה והשוואה ידנית בין שני הגרפים.

## ב. סקיצה של החלופה:



תיאור כללי של החלופה:

חלופה זו מציגה את כל המידע בגרף אחד:

- כל נקודה מייצגת מיזוגות
- מיקום אופקי - רמת ביקוש
- מיקום אנכי - שכר שנתי ממוצע
- צבע הנקודה - שכר שנתי
- גודל הנקודה - רמת ביקוש
- קווים חיצוניים מחלקים את המרחב לארבעה אזוריים

הגרף כולל אינטראקטיביות: סינון לפי רמות ניסיון.

ג. תיאור החלופה במונחים של Marks & Channels

:Marks

Points (עיגולים) -

:Channels

(Quantitative): ביקוש (Position X) -

(Quantitative): שכר שנתי ממוצע (Position Y) -

(redundant encoding): Bikush (Size) -

(Color): שכר (Sequential) -

Text: תווית למיזוגות נבחרות -

Interaction: בחירה, סינון - filtering, selection -

ד. אקספרסיביות ואפקטיביות

#### אקספרסיביות:

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

#### אפקטיביות:

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

#### ה. יתרונות וחסרונות:

##### יתרונות:

- השימוש בערוצו position על שני צירים כמותיים מאפשר השוואת ישירה בין ביקוש לשכר באותו מרחב ויזואלי, ותומך בזיהוי קשרים, דפוסים ויחסים בין משתנים.
- קידוד הביקוש הן באמצעות מקום בציר X והן באמצעות size מספק חיזוק ויזואלי (redundant encoding), המקל על זיהוי מימוניות בעלות ביקוש גבוהה במיוחד.
- קידוד השכר באמצעות מקום בציר Y ובאמצעות צבע בסקירה רציפה מאפשר הבחנה הדרגתית בין רמות שכר שונות, גם כאשר ההבדלים המספריים ביניהן קטנים יחסית.
- קווים החיצוניים מוסיפים מבנה מרחבי לנרטונים ומאפשרים חלוקה סמנטית של המרחב לארבעה אזורים, ללא צורך בהוספה נוספת משתנים חדשים.
- האינטראקטיביות (סינון לפי רמת ניסיון) מאפשרת הפחתת עומס ויזואלי והתקדמות בתתי-אוכלוסיות, בהתאם לעקרון overview → filter → details on demand.

##### חסרונות:

- שימוש במספר ערוצים ויזואליים במקביל (position, size, color, text) מעלה את רמת העומס הויזואלי ודורש ניהול זהיר של תוויות וփיפות.
- ערוצ size נחשב לפחות מדויק מה-position להשוואה כמותית, ולכן אינם מתאימים לביצוע השוואות מדויקות אלא בעיקר לזיהוי יחס סדר וגודל.
- רציפה בין נקודות עלולה להסתיר פריטים מסוימים, ודורשת פתרונות משלימים כגון תוויות סלקטיביות או אינטראקטיביות.

#### 2. הצדקה הבחירה בחלופה הסופית:

החלופה שנבחרה למימוש בשלב האחרון של הפרויקט היא גרפ ה-scatter האינטראקטיבי, המציג את הקשר בין רמת הביקוש למימוניות שונות לבין רמת השכר השנתי הממוצע עבורי. בחירה זו נעשתה לאחר השוואת בין החלופה זו לבין חלופת גרפי ה-Bar, ובהתבסס על התאמת כל אחת מהן למטרות האנליטיות שהוגדרו בשלב הקודם.

חלופת גרפִי ה-`Bar` מאפשרת השוואת מדדיות של כל משתנה בנפרד, באמצעות קידוד של ערכים מסוימים בעורוץ `position` לאוריך ציר משותף, אך אינה מאפשרת ייצוג מושלב של ביקוש ושכר באותו מרחב ויזואלי. כתוצאה לכך, זיהוי קשרים, חריגים ודפוסים בין שני המשתנים מחייב השוואת קוגניטיבית בין גרפים נפרדים ואינו נתמך ישירות על ידי הערכאים הויזואליים.

לעומת זאת, גרפִי ה-`scatter` שנבחר עושים שימוש ב-`position` על שני צירים מסוימים כדי ליצג בו-זמנית את הביקוש ואת השכר, ובכךאפשר ניתוח ישיר של הקשר ביניהם. שילוב ערכיו `color` ו-`size` מחזק את ייצוג המידע ומאפשר זיהוי מהיר של דפוסים וחריגים, כגון מיומנויות בעלות ביקוש גבוה אף ששכר נמוך, או מיומנויות בעלות שכר גבוה ביחס לביקוש. בנוסף, קווים החיצוניים מחלקים את מרחב הנתונים לאזורים סמנטיים ומשיעים בסיווג המיומנויות לקבוצותמשמעותיות.

יתרונו נוסף של החלופה הנבחרת הוא האינטראקטיביות, המאפשרת סינון לפי רמות ניסיון שונות. יכולת זו תומכת בעקרון `pd.options.filter` → `details on demand` → `overview`, ומאפשרת למשתמש לחזור את הנתונים באופן הדרgtiy תוך הימנעות עמוס ויזואלי והתמקדות בתתי-אוכלוסיות רלוונטיות.

למרות שימוש גרפִי ה-`scatter` מורכב יותר ודורש ניהול של עמוס ויזואלי וחיפויות בין נקודות, יתרונותיו האנליטיים והיכולת שלו לתמוך במידה רבה בשימוש מטלות משתמש מצדדים את הבחירה בו כויזואלייזציה הסופית של משימה זו.

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

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

במונחים של התחום: ב\_\_[מקרה 4]\_\_ נבחן כיצד מאפיינים מתקניים ותעסוקתיים משפיעים על רמות השכר בשוק העבודה בתחום ה-AI.

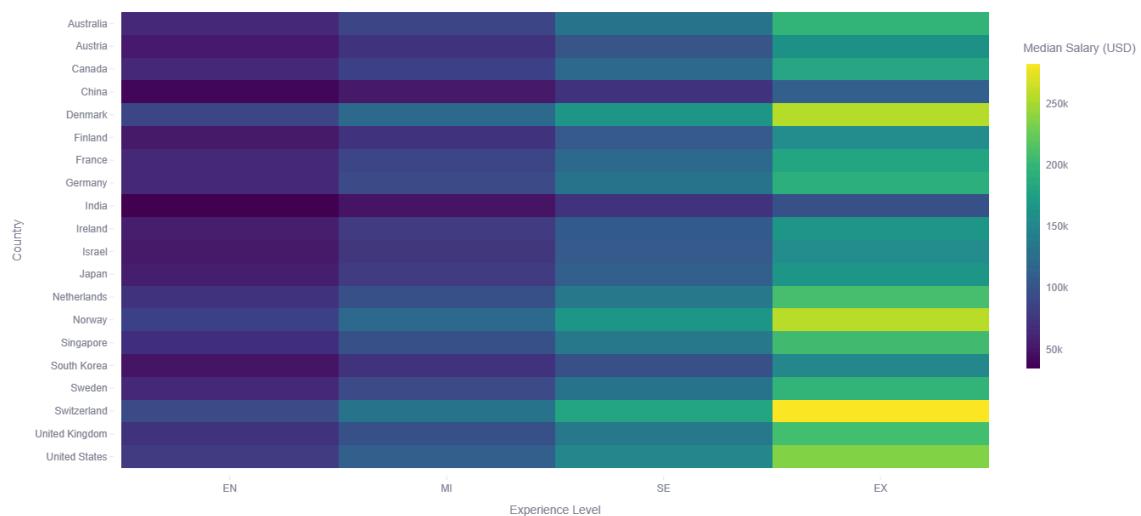
באמצעות ניתוח נתונים שכר מגהגר 2025 *Global AI Job Market and Salary Trends 2025*, נבחן את הקשר בין שכר לבין תפקיד, אזור גיאוגרפי ורמת ניסיון, ונבדוק האם קיימים פערים מובהקים בין קבוצות שונות. במונחים של ana

Action	Sub action	Target	Explanation
analyze → consume	present	correlation	הציגת הקשר בין שכר לבין תפקיד, אזור גיאוגרפי ורמת ניסיון באמצעות זיהוי אמצעות ויזואלייזציות הקשורות
search	locate	outliers	זיהוי משרות או אזורים עם שכר חריג במיוחד (גובה או נמוך)

query	compare	trends	השוואת מגמות שכר בין תפוקידים שונים, אזורים שונים ורמת ניסיון שונות
query	summarize	patterns	זיהוי דפוסים כלליים בשוק העבודה של AI, כגון שילוב תפוקיד אזרור 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 ורמת ניסיון, והשתח

המלבני משמש כ-*container* לערך השכר החיצוני. שימוש ב-area marks מותאים להציג נתונים

בຕבלאות דו-ממדיות שבהן אין צורך להציג פיזור או ערכים בודדים, אלא קשרים בין קטגוריות.

:Channels

- Position (x) – Experience Level (categorical, ordered)

ציר ה-*x* מייצג את רמות הניסיון (EX → EN), סדר שmbossed על היררכיה מושגית קיימת נתונים. מיקום

אופקי, מאפשר קריאה השוואתית בין רמות ניסיון שונות עבור אותה מדינה.

- (al) Position (y) – Country (categorical)

ציר ה-*y* מבידיל בין מדינות. מקום אנסי מאפשר השוואת בין מדינות שונות תוך אותה רמת ניסיון. שילוב שני צירי המיקום יוצר מטריצה שמקסמת השוואות קטגוריאליות.

- (ad) Color luminance / saturation – Median Salary (quantitative, ordered)

הצבע מייצג ערך כמותי רציף (שכר חצוני), באמצעות colormap סדרתי (sequential). שימוש בחישוב luminance/saturation תומך בהבנה בין ערכים נמוכים לגבוהים מבל' להוסף עומס צורני.

#### ד. אקספרסיביות ואפקטיביות

אקספרסיביות:

- כל ערוץ מייצג מידע שקיים בדתא בלבד, אין שימוש בצבע, מיקום או צורה כדי לרמז על משתנה שלא קיים בטבלה המקורית.
- הצבע מייצג משתנה כמותי רציף באופן תקין, השכר החצוני מוצג באמצעות colormap רציף, ללא דיסקרטיזציה מלאכותית (כגון קבוצות "נמרק/גובה"). בכך מנענת יצירת גבולות מודומים שאינם קיימים בנתונים.
- הימנעות מעורצים נוספים שאינם נדרש, אין שימוש בגודל, צורה, צוית או אינימציה - ערוצים שעולים ליצור רמיונות שווא או להעמיס על הפרשנות.
- אין רמייה להיררכיה, סיביות או דירוג הוויזואלייזציה מציגה קשרים והשואות בלבד, מבל' להציג הסבר סיבתי או סדר עדיפות בין מדינות.

אפקטיביות:

- Color luminance מתאים להצגת מגמות ודפוסים כלליים, הצבע מאפשר לזהות ב מהירות אזורים "כהים" ו- "בהירים", וכך לחשוף מגמות רוחביות ובדלים יחסיים.
- סידור מטריציוני תומך בהשואות מרובות:
  - קריאה אנכית: השוואת רמות ניסיון תוך אותה מדינה
  - קריאה אופקית: השוואת מדינות תוך אותה רמת ניסיון מבנה זה ממקסם את האפקטיביות של ערוץ המיקום.
- היעדר ערכים מספריים מפחית עומס קוגניטיבי, המשמש מונחה לקריאה ברמת *overview* ולזיהוי דפוסים, ולא לביצוע *lookup* מדויק שאינו מתרת החלופה.
- מגבלה מודעת של הערוץ הצבעוני, צבע פחות אפקטיבי להשואות מדויקות בין ערכים קרובים, אך מגבלה זו מוצדקת משום שהחלופה אינה מיועדת לדיווק מספרי אלא לניתוח מגמות.

ה. יתרונות וחסרונות:

יתרונות:

- זיהוי מהיר של מגמות רוחביות, לדוגמה: עליה עקבית בשכר עם רמת ניסיון, או פערים ברורים בין מדינות.
- תמייה חזקה במטלות ברמת *overview* החלופה מתאימה במיוחד למינוח *summarize, compare* – *identify patterns*.

- ייצוג קומפקטי של>Data של רב קטגוריאלי, מאפשר הצגת מספר רב של מדיניות ורמות ניסיון בגרף אחד בלבד ליצור עומס צורני.
- ניצול יעיל של ערך המיקום, המטריצה משתמשת במיקום אחד העורצים האפקטיביים ביותר כדי לאפשר השוואות רבות.

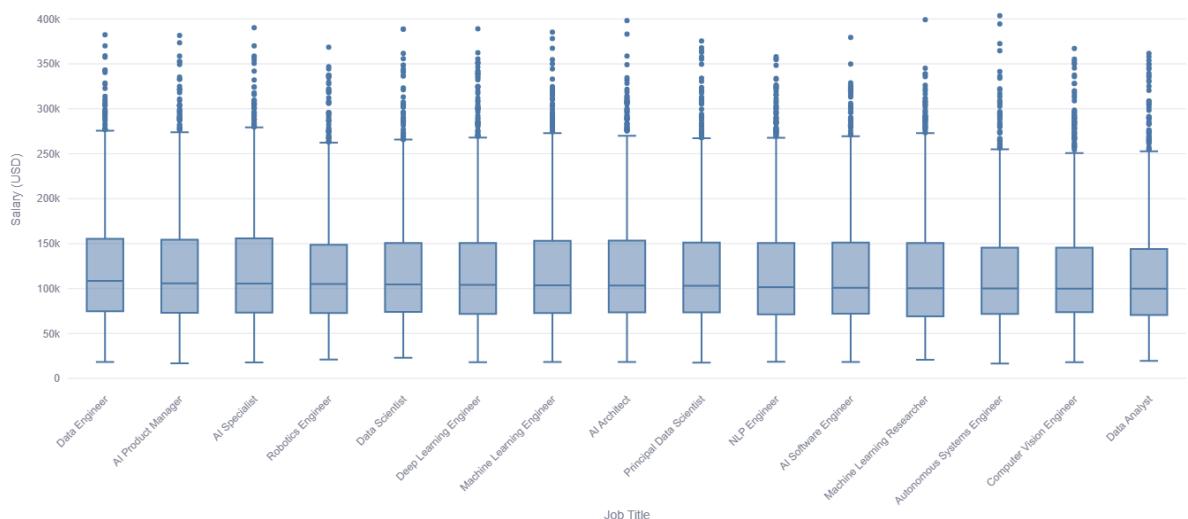
**חרונונות:**

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

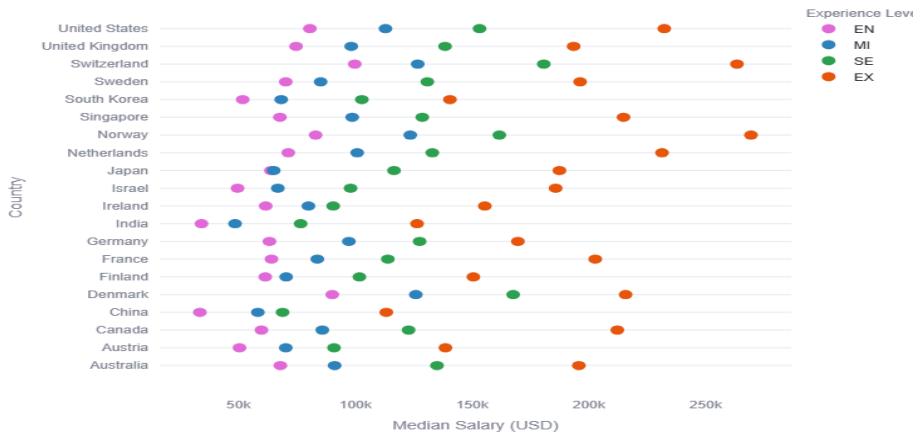
#### ביתוח החלופה השנייה:

ב. סקיצה של החלופה:

Boxplot – Salary Distribution by Job Title .1



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



ג. תיאור החלופה במונחים של :Marks & Channels

Graph 1: Boxplot – Salary Distribution by Job Title

:Marks

- Area marks boxes - כל תיבת מייצגת את ה- $25\%-75\%$  IQR של השכר עבור תפקיד מסוים. זה mark שמשתמש להראות שונות פנימית ולא רק "מספר אחד" לכל תפקיד.
- Line marks median + whiskers - הקוו בתוך התיבה הוא החציון, וה-whisker מציגים טווח טיפוסי. בערך מתקובל סיכון חזק של ההתפלגות שמאפשר להשווות "מרכז" מול "פיזור" בין תפקידים.
- Point marks outliers נקודות החיריגים מצביעות על משכורות נדירות/קיצוניתות שאינן חלק מהטוויה הטיפוסי. זה תומך במשמעות של איתור חיריגים (locate) בלי להסתיר אותם בתוך הסיכום.

:Channels

- Position (x) – Job Title (categorical, unordered) המיקום האופקי מבידיל בין קטגוריות תפקיד. לאחר שאין סדר טבעי לפקידים, המיקום שביצעת לפניו הוא החלטת עיצוב שמשרתת השוואה.
- Position (y) – Salary (quantitative) ציר ה- $y$  מאפשר השוואת שיער כי שכר בין תפקידים, משום שמייקום הוא העורך התפיסתי המדיוקן ביותר להשוואה כמותית.
- Color (single hue) – neutral / no additional encoding כאן צבע לא מקודד משתנה מסוים, אלא משמש כאחדות ויזואלית. זה מפחית עומס ומונע פרשנות שגוייה ("אם צבע שונה – יש משמעות").

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

:Marks

- Point marks - כל נקודה מייצגת חcyון שכר עבור צירוף של מדינה ורמת ניסיון. זה mark "דוחה" שמאפשר להציג הרבה קבוצות בלבד להעמס כmo `dotplot` לכל קבוצה.

:Channels

- Position (x) – Median Salary (quantitative

שוב נבחר המיקום, כי המטרה היא השוואות בין חcyוניים: מי גבוהה יותר, כמה פער יש, ועוד. זה העורך המדיוק ביותר להשואות כמותיות, ולכן מתאים מאוד לשימוש `compare`.

- (Position) 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 "סרייה" אופקית.

- Color hue – Experience Level (categorical, conceptually ordered

צבע בגון (hue) מתאים לקטגוריות, כי המטרה היא להבדיל בין EX/SE/EN/MI/EN ולא למדוד גודל כמותי. לעומת זאת שיש סדר מסווגי בין הרמות, ההשוואה הכמותית עדין נעשית באמצעות x ולא באמצעות הצבע וזה מונע הטעיה.

- Size – constant

הגודל קבוע כדי לא "להציג" ממשועות שלא קיימת בנתונים. זה מצבם סיכון של פירוש שגווי (למשל לחושב שגודל = מספר משרות).

ד. אקספרסיביות ואפקטיביות

אקספרסיביות:

- כל עורך מייצג רק מידע שקיים בנתונים, אין שימוש בצבע/גודל/צורה כדי "לرمז" על משתנה שלא מופיע בטבלה. כך נמנעת הטעיה סמנטית.
- אין קידוד כפול של אותו משתנה, השכר מוצג כמותית באמצעות מיקום הצבע ב-`dot` מקובד רק ניסיון ולא את השכר שוב.
- Boxplot – צבע ניטרלי כדי לא ליזור ממשועות נוספת, שימוש בגוון אחיד מונע מצב שבו המשתמש חושב שהצבע מייצג קטgorיה/aicوت/קבוצה נוספת ללא קיימת.
- Dot plot – צבע מייצג משתנה קטgorיאלי מתאים (experience level), זה מיפוי "נק". אין ניסיון להשתמש בצבע כדי למדוד "כמה" אלא כדי להבדיל בין קבוצות.
- הימנעות מערוצים שעולים להטבות, לא נעשה שימוש בשטח/נפח/גדלים משתנים ללא ממשועות, וכך נוצרת היררכיה חזותית מזויפת.

אפקטיביות:

- Position לערכים כמותיים = השוואת מדיקת יותר, בשני הגרפים השכר ממופה למיקום, ולכן קל לבצע `compare` בין תפניות/מדינות/רמות ניסיון בצורה מדיקת יחסית.

- Color משמש להפרדה קטגוריאלית ולא למדידה, ב-Dot plot הצבע נועד לזהות את קבוצת הניסיון.
- ההשוואה של "מי גובה יותר" נשענת על מיקום, מה שמעלה דיוק תפיסתי.
- פיצול לשני גרפים תומך במיקוד מטלות, במקרה "לדוחוף הכל" לגרף אחד, כל גרף ממוקם אפקטיבית 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 התפלגות.
- Dot plot אפקטיבי ל-*summary comparisons* בקנה מידה גדול, הוא מאפשר להשוות הרבה מדינות ורמות ניסיון בו-זמנית, תוך שימוש על גרף "דחוס" שלא מתפוץץ בעומס marks.
- הצדקהiae לשימוש במספרים על הגרף, העומס ויזואלי בלי לשרת את המשימה המרכזית.

#### ה. יתרונות וחסרונות:

יתרונות:

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

חסרונות:

- Boxplot פחות נוח להשוואה מספרית "מדויקות" של חיצונים, אפשר לראות מי גובה יותר, אבל קשה להוציא ערך מדויק בלי hover/טבלה - זה לא לטוב *trade-off* overview.
- Dot plot לא מציג פיזור או שונות פנימית, הוא מראה רק חיצון, ולכן שני מקומות עם אותו חיצון יכולים להיות בעלי התפלגות שונה לגמרי - מידע שאבוד.
- ריבוי קטגוריות (מדינות) דורש יותר קשב/גיליה, כישר הרבה מדינות, המשתמש צריך לסרוק לאורך ציר על זה מגדייל מאמץ תפיסתי בהשואות רחבות.
- תלות בבחירה צבעים ב-dot plot, אם הגוונים קרובים מדי או לא מותאים לעיורן צבעים, הבדיקה בין קטגוריות נפגעת. לכן ח"בים palette מובהן ויציב.

#### 2. הצדקת הבחירה בחלוקת הסופית:

חלוקת השנייה נבחרה למיושן הסופי משום שהיא מספקת מענה רחב ומדויק יותר למטלות שהוגדרו, תוך שימוש של עקרונות האקספרסיביות והאפקטיביות של Munzner. השילוב בין *dot plot* להṯפלגות שכר לפי תפקיד לבין *dot plot* המציג שכר חיצוני לפי מדינה ורמת ניסיון מאפשר ניתוח רב-ממדית של הנתונים מבלתי לדחוס משתנים רבים לגרף יחיד. חלקה זו תומכת של השוואה ויזיהו מגמות והן בזיהוי שונות פנימית וערכית חריגים, שאינם ניתנים ליזיהוי ברור בחלוקת הראשונה. בנוסף, השימוש בערכים ויזואליים אפקטיביים - מיקום להשואות כמותיות וצבע להבחנה קטגוריאלית - משפר את הדיוק התפיסתי ומפחית עומס קוגניטיבי. לפיכך,חלוקת השנייה

מאפשרת עומק אנליטי גבוה יותר לצד קריאות חזותית טובה, ומרקפת התאמת מובהקת יותר למטרות הניתנות ולמודל התכנון של Munzner.

#### **חלק 4: הסבר על היצוגים ויישומים:**

לפני הначלה העבודה עם המידע נדרש ביצוע Pre-processing ראשוני על כל הדאטה שכלל חיבור שני קבצי המידע. עשינו זאת באופן הבא: המידע במקור הגיע לאתר 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 הממחקר, ולאחר מכן איחוד הדאטאות תוך תיקון שדה "id\_job" בקובץ שהעבכנו כך ש  $new\_job\_id = old\_job + 15,000$  כדי לאפשר מזהה ייחודי לכל רשותה.   
כזאת ייעזר בהמשך את הדashboard יש להוריד את העדכון החינמי של streamlit לגרסה 1.35.0, עדכן זה היבנו חינמי!

#### **מקרה 1: האם קיימים הבדלים גיאוגרפיים מובהקים באימוץ AI ברחבי העולם?**

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

#### **מסמך 1: Comparator**

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

1. הנדסת מאפיינים - זיהוי מקור עובדים: יצרנו משתנה בוליאני חדש בשם `is_same_location`. משתנה זה מחושב על ידי השוואתו בין עמודת `employee_residence` (מקום העבודה) לבין `company_location` (מקום העבודה) (מקום החברה). פעולה זו מאפשרת לנו לפחות בזמן אמת את כוח העבודה ל"מקום" (מתוך המדינה) מול "בינלאומי/מקום חוץ" (מחוץ למדינה), נתן קריטי להבנת דפוסי העסקה גלובליים.
2. ארגזציה לטיפול בערכים חסרים: כדי ליצור את גרפף "סולם השכר" (Salary Comparison) ולהשווות אותו למומוצע העולמי, חישבנו מראש את ממוצעו השכרי הארציים לפי שנות ניסיון (`avg_global_exp`). חישוב מקדים זה מאפשר לנו להציג קו "יחסו יroke" ("Global Avg") בכל תרשימים, גם כאשר בוחרים מדינה שיש לה נתונים חלקיים בלבד עבור שנות ניסיון מסוימות.
3. המרה וטיבוב: ביצענו המרת הנתונים בוליאניים למספרים (Integer) כדי לאפשר ספירה וחישוב אחוזים מדויק עבור תרשימים הבר האופקי המציג את התפלגות מקור העובדים.

ב. שינויים במימוש הוויזואלייזציה לעומת התכנון המקורי במהלך הפיתוח, נדרשו לבצע התאמות כדי לשפר את חווית המשתמש בדASHBOARD הראשי:

1. שיפור הקוד הוייזואלי:
  - קנה מידת דינמי ביציר ה-Z: בתכנון המקורי, גרפף התפקידים המוביילים ("Top Roles") התחיל תמיד מאפס. נוכחנו לדעת כי במידינות מסוימות, הפרערם בין מספר השירות 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. שניי זה מדגיש את ההבדלים הווייזואליים גם בפערים קטנים.

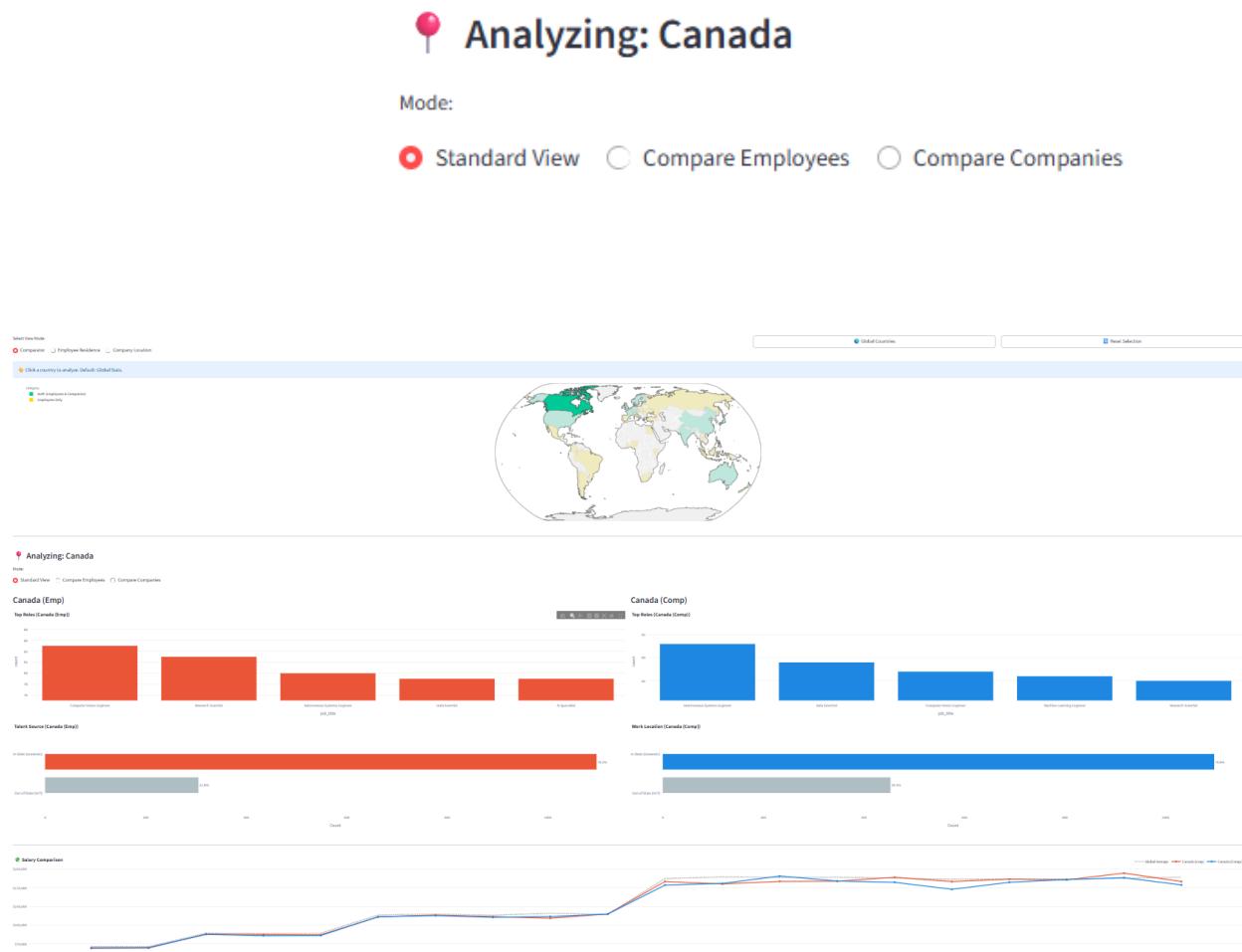
- מנגנון השוואה (Comparator Mode): במקור תוכננו להציג רק נתונים עבור מדינה בודדת. תוך כדי מימוש, הבנו שהערך האמתי נוצר מהשוואה. לכן, פיתחנו את מצב "Comparator", המשמש במקרה לא רק לבחירה בודדת אלא לבחירה כפולה (מדינה ראשית ומשנית), ומחליף את התצוגה הגלובלית בתצוגה מופצת המעתה נתונים של שתי מדינות זו לצד זו.
- כטור Global Countries: שילבנו כטור ייעודי בראש הממשק המאפשר גישה מהירה לנtíוני מסוימים. במבט זה, לחיצה על הכטור 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 כנתוניים גלובליים.
- 2. ניהול מצב (State Management): המעבר בין תצוגת "מגורוי עובדים" לתצוגת "מיקום חברות" דרש שניי אריכטוני. במקומות לטען מחדש את הדף ולאבד את המדינה שנבחרה, הטמענו שימוש ב-session\_state כדי "לזכור" את המדינה הראשית והמשנית גם בעת החלפת מצב תצוגה או רענון רכיבים.

## ג. אופן המימוש תמונה 1:



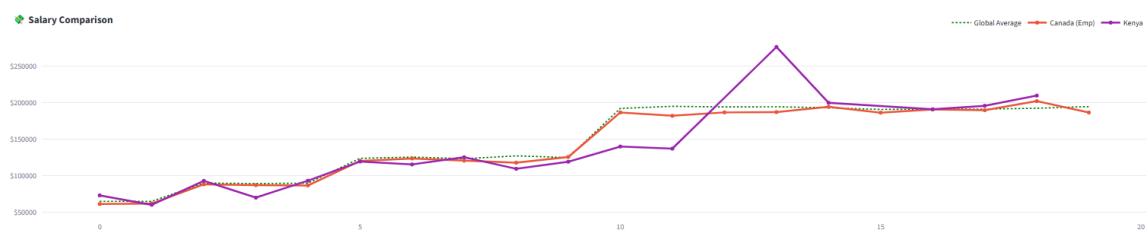
**מסך הבית - מבט גלובלי (Global Landscape)** תצלום זה מציג את מצב ברירת המחדל. המפה האינטראקטיבית צבועה לפי סוג הפעולות (עובדים/חברות/גמ' וגמ'), ומתחתייה מוצגים ארבעה גרפים (תפקידים, מקור עובדים, שכר לפי ניסיון, וסולם שכר) עבור כל העולם.

תמונה 2: אינטראקטיבית - Drill-down למדינה (Single Country View) כאשר המשתמש לוחץ על מדינה במפה (למשל, קנדה), הקוד מזיהה 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 האגוצבלי (הקו הירוק המקבוקו) להשוואה.



תמונה 3: מצב השוואה (Comparator Mode) במצב זה, המשתמש בוחר שתי מדינות שונות במפה. המסל מתפצל לשתי עמודות: עמודה ימנית למדינה הראשית ועמודה שמאלית למדינה המשנית. בתחתית המסך נוצר גרפ קוו "יעודי" (create\_comparison\_line\_chart) המציג את עקומות השכר של שתי המדינות שנבחרו על גבי מערכת צירים אחד, ומאפשר זיהוי מהיר של פערים כלכליים.





ד. כתיבת הקוד: בקוד שנכתב עברו ויזואלייזציה זו נעשה שימוש בספריות ובכלים הבאים:

- הציגות הגלובלית והן עברו התוצאות המפוצלות במצב ההשווה. מודולריות סביב פונקציית `create_graphs`, גישה שאפשרה שימוש חוזר באותו קוד בדיק – הן עברו בספריה אפשר להعبر את `DataFrame` כקלט לפונקציות גנריות המיצירות את הגրפים. `Pandas`: לטובות עיבוד וניהול הנתונים, החל מטעינת המידע ועד להכנתו להצגה ויזואלית. השימוש באיפוס הבחירה בעת שינוי מסננים או מעבר בין מצב תצוגה.

Streamlit: עברו בניית ממשק Frontend (צד לקוח) וניהול לוגיקת האפליקציה. נעשה שימוש נרחב ב-`session_state`. st לצורך ניהול מורכב של בחירת מדיניות (Secondary ו-Primary) ולמנועת-

#### מוצר 2: מגוריו 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 לזרזות דפוסי הגירה דיגיטלי:

1. פילוח גאוגרפי: הנתונים עברו קיבוץ (Grouping) לפי עמודת employee\_residence. פעולה זו מאפשרת לנו להתייחס למיקום הפיזי של העובד כנקודות המוצא לניתוח, ללא קשר למיקום החברה המפעילה.
  2. הנדסת מאפיינים - ניתוח מקור התעסוקה: השימוש במשתנה המחשב chosen\_location\_is מוביל כאן למשמעות ייחודית. במבט זה, הנתון משמש להציג אחזות העובדים במדינה מסוימת המועסקים על ידי חברות她们 (מיקור חוץ/Remote) לעומת חברות מקומיות, מה שמעיד על יכולת הtalent המקומי והביקוש הגלובלי אליו.

ב. שינויים בימוש הויזואליות לעומת התכונן המקורי במהלך היפויו ביצעו התאמות כדי להבדיל ויזואלית בין מסר זה למסר החברים:

1. שפה חוזית: המבט מאופיין בשימוש עיקרי בצבע אדום. בחירה זו מאפשרת למשתמש להבין אינטואיטיבית שהוא צופה בנטוינו עובדים, ומונעת בלבול מעבר בין המ██ים.
  2. המפה: מציגה מפת חום (Choropleth) של ריכוז עובדים. סקלאלת הצבעים מנורמלת ביחס למקסימום הגלובלי (GLOBAL\_MAX\_COUNT), כך שמדיניות עם ריכוז גבוה של מומחי AI בולטות בכהותן באופן פרופורציונלי.

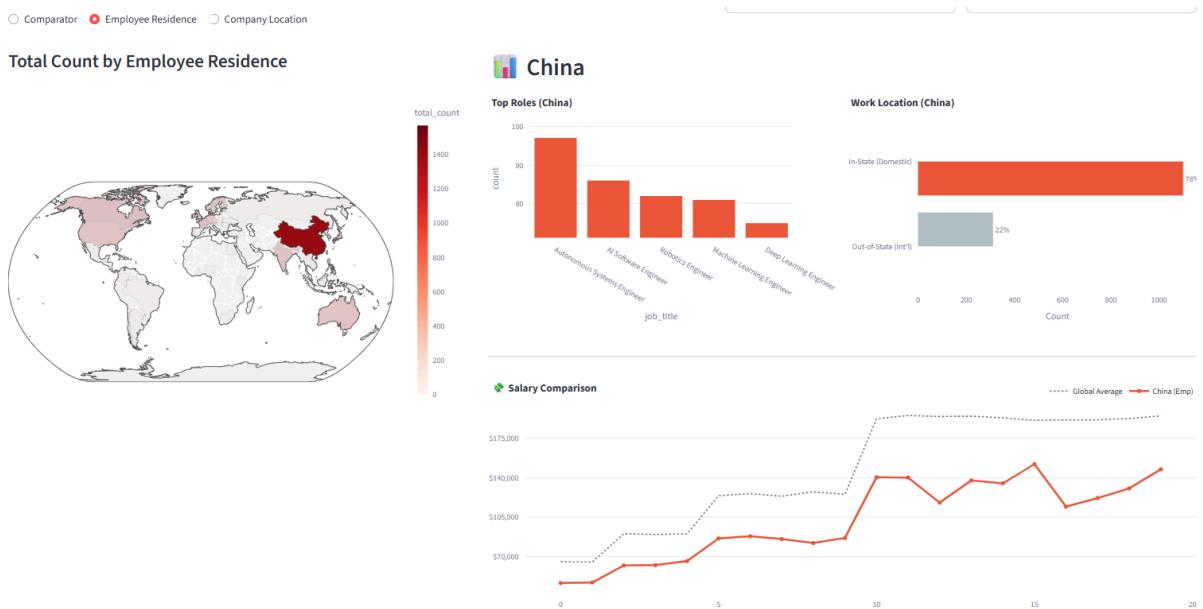
3. כפטור Global Countries: שילבנו כפטור ייודי בראש הממשק המאפשר גישה מהירה לנוטרי מאקרו. במבט זה, לחיצה על הכפטור חושפת טבלה המדרגת את 10 המדינות המובילות במספר העובדים. פונקציה זו מאפשרת למשתמש לzechות במיהירות את מוקדי הכוח של הטalent העולמי מבל' לסרוק את המפה כולה. בנוסף, במצב דיפולטי אנו מציגים את הנתונים של מדינות אלה כתונונים 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: גրף ברים אופקי הממחיש כמה מההתושבים עובדים "פנימה" (לשוק המקומי) וכמה "חוצה" (לחברות בינלאומיות).
- Salary Comparison: משווה את עקומת השכר של התושבים ממוצע הגלובלי (קו יירוק מקווקו).
- Talent Source: משווה בין כמה אחוז מהעובדים עובדים בחברות במדינה הנבחרת לעומת כלל השוכדים במדינות אחרות מהרשימה

## ג. אופן המימון

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



## טרכן 3: מיקום חברות (Company Location View)

א. שינוי preprocessing למורota שהמבנה הכללי זהה למסך הקודם (לצורך עקביות), הלוגיקה והתוכן משתנים כדי לשקף את צד החברות:

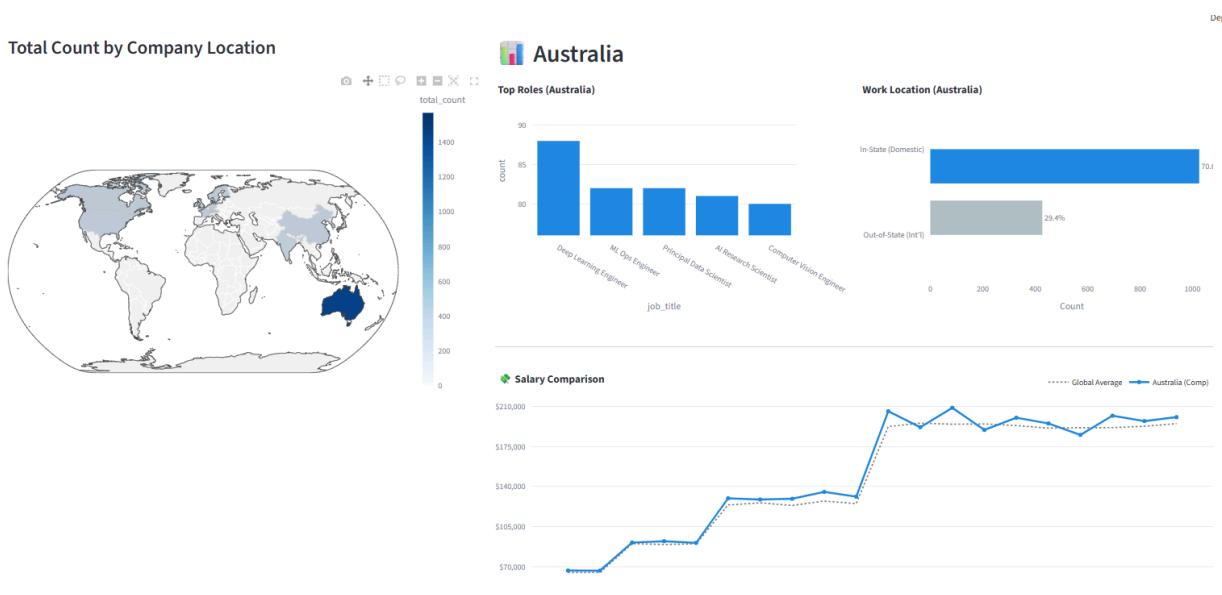
- .company\_location
1. מקור הנתונים: הסיכון והאגרגציה מתבצעים הפעם לפי עמודת `company_location`
2. משמעות המבדדים: הגרף המציג "מוקמי מול זר" מקבל משמעות הפוכה – הוא מראה כמה מכוח העבודה של חברות במדינה זו הוא מקומי, וכמה גויס מדיניות אחרות (Offshore/Remote).
3. ציר השכר: מייצג את "נכונות לשלם" (Willingness to Pay) של חברות באזורי זה, ולא בהכרח את יוקר המחייה של העובדים.

ב. שינויים במימוש הווייזואלייזציה לעומת התכנון המקורי התייחסות העיקריים נועדו לחדר את ההבדלים בין מדינות עם שוקים צפופים:

1. שפה חזותית: המבט מאופיין בשימוש עקבי בצבע כחול שני הצביע הוא קרייטי למניעת עומס קוגניטיבי בעת המעבר בין הממצבים.
2. דינמיות בציריהם: בדומה למבט הראשון, גם כאן ישם מנגן לחישוב דינמי של ציר ה-`Y` בגרפים (\* min 95, 0). מנגן זה פותח כדי להבליט פערים קטנים בין פופולריות של חברות השונות, שלא היו נראים בגרף המקורי מAffected.
3. כפטור Global Countries: הcpfutor משנה את תפקידו בהתאם למבט. במבט זה, לחיצה על הכפטור חושפת טבלה המדרגת את 10 המדינות המובילות במספר השירותים (חברות). פונקציה זו מאפשרת למשתמש לראות היכן מרכז ההון והיכן נמצאים Tech Hubs הגדולים. בנוסף, במצב דיפולטי אנו מציגים את הנתונים של מדינות אלה כבסיס להשוואה.

#### ג. אופן המימון

תמונה 1: מבט Company Location בתצלום זה המשמש בחר **באוסטרליה**. ניתן לראות כיצד מפת העולם והגרפים (צבע כחול) משקפים כתעת נתוניים מנקודת המבט של המעסיקים האוסטרליים, כולל נוכנותם לשלם עבור טאלנטים בינלאומיים.



#### ד. כתיבת הקוד:(משותף למסכים 2 ו-3)

- בקוד שנכתב עבור שני המבטים הללו , נעשה שימוש בספריות ובכליים הבאים:
- Streamlit: שימשה לבניית משיק המשתמש ולניהול הniest בין המבטים (באמצעות רכיב `radio`).
  - Pandas: שימשה לטעינת קובץ CSV ולביצוע מניפולציות על הנתונים (כגון Filtering-Grouping).
  - Plotly Express & Graph Objects: שימשו לייצרת הויזואלייזציות האינטראקטיביות. הבחירה ב-`Plotly` מאפשרת לנו להטמיע יכולות Zoom ו-Hover מתקדמות בשני המבטים.
  - שימוש חוזר בקוד (Code Reusability): כדי לשמר על קוד נקי ויעיל, נמנעו משכפל קוד. במקום זאת, כתבנו פונקציה גנרטית אחת בשם `create_graphs`. פונקציה זו מקבלת כפרמטרים את כל DataFrame המשווני, הכוורת הרצiosa וצבע הנושא (אדום או כחול). כך, החלפת מבט מעדכנת את כל השימוש ללא טעינת דף מחדש.
  - ניהול מצב (State Management): אתגר מרכזי היה לשמר על בחירת המדינה בעת המעבר בין המבטים. השתמשנו ב-`state.session`. `st` כדי לשמר את המשתנה `primary_country` כדי ביכולת כך, האינטראקטיה היא רציפה: המשתמש יכול לבחור מדינה במבט אחד, ולאחר מכן במבט השני כדי לראות מיד את הנתונים המשלימים עבור אותה מדינה.
  - שימוש בכל עזר (AI): במהלך הפיתוח נעשה שימוש במודול 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.

#### מטרה 2: ניתוח השוואתי רב מימדי של תפקידי עבודה בתחום AI.

שימוש מקדים נוסף: בשלב העיבוד המוקדם הנתונים נתונים מקובץ CSV ובערו ארגזיה לפי `title_job`, כך שכל תפקיד מיוצג בערך אחד לכל ממד. עבור ממדים מסוימים (כגון שכר, ניסיון, יחס עבודה מרוחק והטבות) חושב ערך ממוצע, ועבור ממד הביקוש חושבה ספירה של מספר המשרות. מאחר שהמדדים נמדדים בסקלאות שונות, בוצע נירמול מסוג Min-Max לכל ממד בנפרד לטווח [0,1], על מנת לאפשר השוואה ויזואלית בין ממדים הטרוגניים. במקביל נשמרו הערכים הגלומיים, אשר מוצגים באינטראקטיבית `hover` ובטבלת נתונים משלימה. בנוסף, הוחל ערך מינימלי ויזואלי לצורכי תצוגה, המאפשר אינטראקטיה גם עבור ערכים מנוימים נומקיים במילוי. תצוגה דיפולית:

## 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

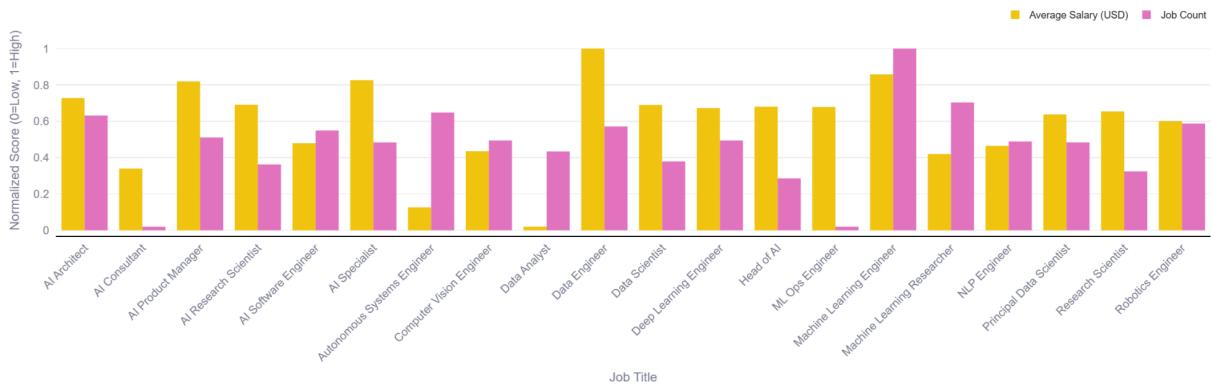
Average Salary (USD)  Job Count  Years of Experience  Benefits Score  Remote Work Ratio (%)

### 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 קרייטריונים להשוואה ניתן לסמן אותם בסרגל

ניתן לבחור בין אחד לחמשה פיצ'רים שאוטם אנחנו רוצים להשוואות ולראות את השוני בינהם

#### Select Metrics for Comparison

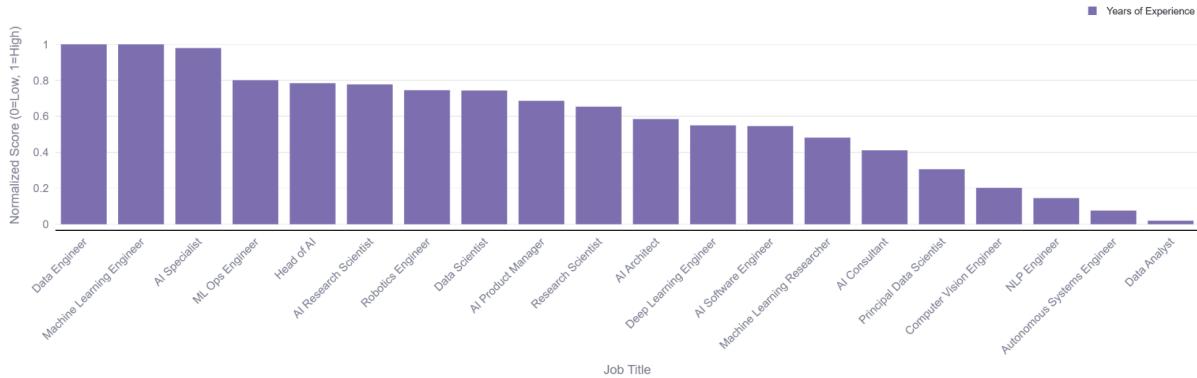
 Average Salary (USD) Job Count Years of Experience Benefits Score Remote Work Ratio (%)

#### Chart Sort Settings

Sort Chart By:

Years of Experience

#### Normalized Multidimensional Comparison (Sorted by: Years of Experience)



#### Select Metrics for Comparison

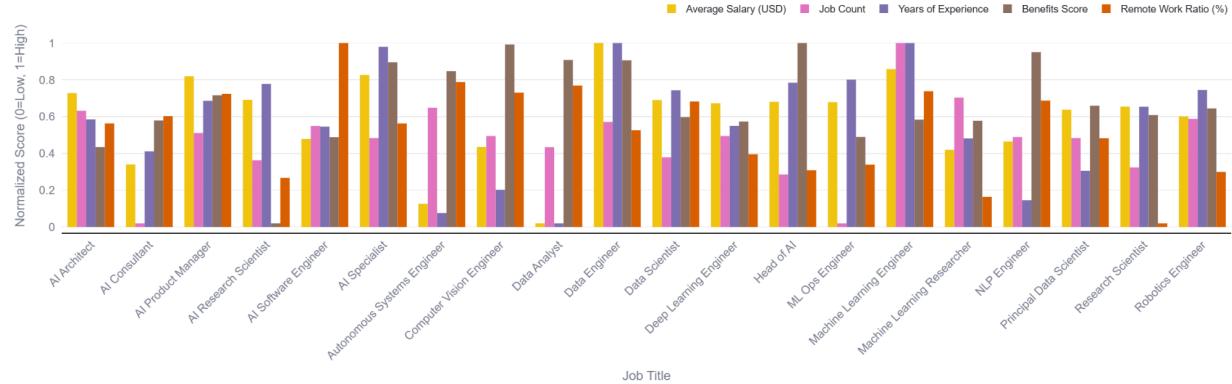
 Average Salary (USD) Job Count Years of Experience Benefits Score Remote Work Ratio (%)

#### Chart Sort Settings

Sort Chart By:

Name (A-Z)

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



בנוסף ניתן לעמוד על עמודה רצiosa ולקבל את הערך המקורי שלה ואת הערך היחסי לאחר הנרמול  
וכן למיין את סדר העבודות לפי פיצ'רים מסוימים לבחירה

### Select Metrics for Comparison

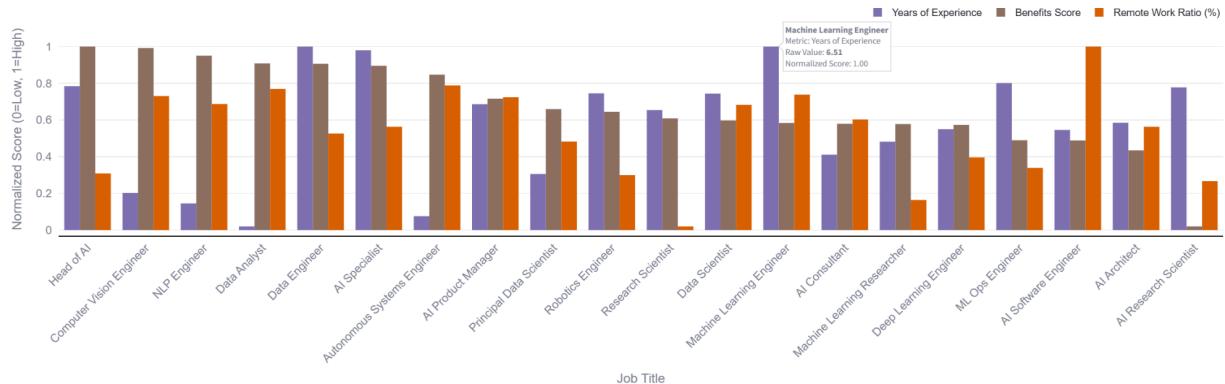
Average Salary (USD)  Job Count  Years of Experience  Benefits Score  Remote Work Ratio (%)

### 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 למשלה. הערכים בטבלה הם הממצאים של כל שילוב ולכל עמודה יש מפת חום - ככל שהצבע יותר כהה כהה 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 סוג העבודה והעמודות מייצגות את הפיצרים שנבחרו למשלה

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

## Underlying Data (Raw Values)

 Search by Job Title

All

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 - לטיעת הנתונים, עיבוד ואגרציה לפי תפקיד העבודה (חישוב ממוצעים וספירות).
- Streamlit - לבניית ממשק משתמש אינטראקטיבי הכלול בחירת מדדים, הצגת גרפים וטבלאות.
- Plotly Express - ליצירת גרפי נתונים אינטראקטיביים עם תמייהה-over והשוואה בין מדדים.

הקוד מומッシュ אפליקציית Streamlit ומחולק למספר שלבים לוגיים בrhoים. תחילת מוגדר מבנה העמוד והcotrols, ולאחר מכן נטען הקובץ ומבצעת בדיקת תקינותו לקיומו.

משחק המסתמך לאפשר בחירה דינמית של מדדים באמצעות תיבות סימון, כאשר כל שינוי בבחירה גורם לרענון מיידי של הוויזואלייזציה. בחירה זו משפיעה הן על הנתונים המוצגים והן על עיצוב הגרף (כגון מספר העמודות, צבע והמרקא).

הגרף עצמו מומחש כתרשים עמודות מקובץ (Grouped Bar Chart), אשר לכל מדד מוקצת צבע קבוע ועקבי. הערכים המוצגים בגרף הם ערכים מנורמליים, אך בעת מעבר עם העכבר על כל עמודה מוצג חלון מידע (hover) הכלול את שם התפקיד, שם המدد, הערך הגלומי והערך המנורמל.

בנוסף לgraf, הקוד מציג טבלת נתונים משלימה המציג את הערכים הגלומיים של כל המדרדים לכל תפקיד, עם קידוד צבעים רציף. טבלה זו מאפשרת אימיות של המציג בgraf והעומקה בניתוח, ללא צורך באינטראקטיבית עם התרשים עצמו.

לבסוף, הוגדרו מאפייני עיצוב כלליים (רקע לבן, סידור מקרא, סיבוב תוויות ציר) על מנת לשפר את קריינות הוויזואליות ולהתאים להצגה בדו"ח או במצבת.

### מטרה 3: השוואת בין רמת הביקוש למיומנויות שונות בשוק העבודה בתחום ה-AI לבין רמת השכר המוצעת בעברן.

א. ביצוע Pre-processing:

במסגרת יצירת הוויזואליות בוצע תהליך עיבוד מקדים על בסיס הנתונים הגלומיים כדי להתאים לניתוח סטטיסטי ולהציג ויזואלית זו. הנתונים המקוריים הגיעו ברמת "משרה בודדת", בעוד שמטרת הנתיחה היא בחינת "מיומנויות טכנולוגיות".

להלן שלבי העיבוד שבוצעו:

1. פירוק מחרוזות כישורים: המידע בעמודות required\_skills הופיע במקור כמחרוזת טקסט אחת המכילה מספר כישורים מופרדים בפסיקים (למשל: "Python, SQL, AWS"). בוצעה פעולה המירה שהפכה כל מחרוזת כזו לרשימה של כישורים נפרדים.
2. שינוי מבנה הנתונים: בוצעה פעולה explode על רשימות הכישורים. פעולה זו שינה את ייחדות הנתיחות הבסיסית של הטבלה: משורה המייצגת משרה, לשורות המייצגות כישור בודד. המשמעות היא שאם משרה אחת דרש 5 כישורים, היא פוצלה ל-5 שורות נפרדות, כאשר נתונים השכר והnisian שכפלו לכל שורה. פעולה זו הכרחית כדי לחשב סטטיסטיות עבור כל כישור בנפרד.
3. טיב וקיצור שמות: בוצע נרמול של שמות המיומנויות כדי לשפר את קרינות הגרפ. שמות ארכיכים הוחלפו בKİזרים מקובלים בתעשייה (לדוגמה: "Natural Language Processing" הומר ל-"NLP", ו-"Machine Learning" ל-"ML").
4. אגgregציה וחישוב מדדים: הנתונים קובצו לפי סוג המיומנות, וחושבו שני מדדים מרכזיים לכל כישור:
  - Average\_Salary
  - Job\_Count
5. סינון רעים סטטיסטיים: בוצע סינון להסרת מיומנויות נדירות המופיעות פחות מ-10 פעמים במאגר הנתונים. סינון זה נועד להבטיח מובהקות סטטיסטית ולמנוע הטוויות הנובעות ממדגם קטן מדי ("outliers").
6. הנדסת מאפיינים לתצוגה: נוצר מדד מחושב חדש בשם Score, המשלב את השכר והביקורת המנורמלים. מדד זה שימש לקביעת סדרי העדיפויות להציג תוויות הטקסט בגרפ, על מנת למנוע עומס ויזואלי ולהציג רק את המיומנויות המשמעותיות ביותר.

ב. שיפור המימוש הויזואלי:

במהלך תהליכי הפיתוח והשימוש, בוצעו מספר התאמות לעיצוב המקורי. שינויים אלו נבעו מהתובנות שעלו תוך כדי עבודה עם הנתונים האמתיים, ומtower נדרש לשפר את קרינות הגרפ ואת חווית המשתמש.  
להלן השינויים המרכזיים שבוצעו:

## 1. שיפור הקוד הויזואלי:

- סקאלת צבעים דיפרנציאלית: במקומות צבע אחד או אקריאי, הוגדרה סקאלת צבעים רציפה המיצגת את גובה השכר. הסקאללה מדגישה בבירור את הקצוות (שכר גבוה מאוד מול נמוך מאוד) ומציגת צבע ניטרלי עבור ערכי הממוצע, מה שמאפשר זיהוי מגמות ממבט מהיר.
- הוספה מסגרות: נוספו קווים 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 קווים ייחודיים: בניגוד לתכנון הראשוני שעשי היה להציג רק ענן נקודות, נוספו למימוש הסופי קווים אנכיים ואופקיים המציגים את החזיניות של השכר והביקורת. קווים אלו מחלקים את הגרפף לאربעה רביעים ברורים, מה שמעניק למשתמש הקשר מיידי ומאפשר לו להבין האם מיפוי מסוימת נמצאת "מעל" או " מתחת" לשוק הממוצע.
- 3. מניעת עמוס ויזואלי בתוויות: בשל צפיפות הנתונים, זהותה בעיה של חפיפת טקסטים שהקשטה על קריית שמות המיפויים. במימוש הסופי פותחה לוגיקה דינמית למיקום התוויות, המזינה את הטקסט ביחס לנקודה בהתאם למיקומה למרחב ולצפיפות הסביבתיות, כך שהשימוש נשארים קריאים וברורים.
- 4. אופטימיזציה של חווית הבחירה: במשק הסיכון שונתה ברירת המחדל כך שכל רמות הניסיון מוצגות ונבחנות אוטומטית עם טעינת הדף. גישה זו נבחרה כדי להעניק למשתמש תמורה מצב מלאה ורחבה כבר בשלב הראשון, תוך שמירה על הגמישות להסיר או לבחור מחדש רמות ספציפיות בקלות.

## ג. אופן המימוש:

הסבר כללי על הויזואלייזציה:

להלן פירוט המרכיבים הויזואליים והלוגיקה שמאחוריהם:

### 1. מערכת הצירים:

- ציר ה-X (ביקוש): מייצג את ה-Job Count, כלומר מספר המשרות במאגר הדוחשות את המיפויים הספציפיים. ציר זה מעיד על הפופולריות והדרישה למיפויים בשוק.
- ציר ה-Y (שכר): מייצג את ה-Average Salary, כלומר השכר השנתי הממוצע (ב דולרים) המוצע למשרות הדוחשות מיפויים זו.

### 2. קידוד ויזואלי רב-מדדי: כדי להעיר את המידע המוצג בכל נקודה בגרפף, נעשה שימוש בשני ערכאים ויזואליים נוספים מעבר למיקום בציריהם:

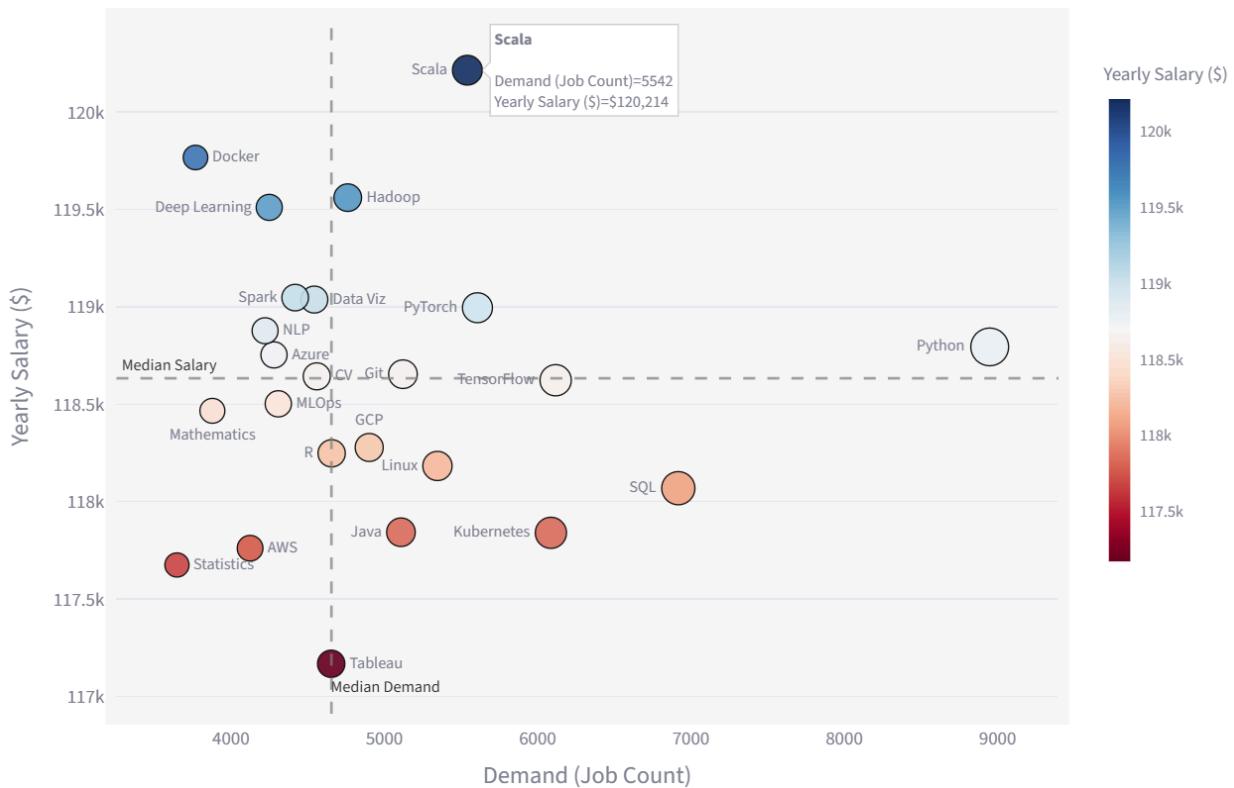
- גודל הנקודה: גודל הבועה מושפע באופן ישיר מהביקוש (Job Count). בוועות גדולות יותר מייצגות מיפויים פופולריים יותר. בחירה זו נועדה למשוך את עין המשתמש באופן מיידי למיפויים הדומיננטיות בשוק.
- צבע: נעשה שימוש בסקאלת צבעים רציפה המציגת את גובה השכר. קידוד זה מספק אינדיקציה כפולה לשכר (גם המיקום בציר ה-Y וגם הצבע), מה שמקל על>Zיהוי חריגים (למשל, מיפויים עם ביקוש נמוך אך צבע המעיד על שכר גבוה מאוד).

### 3. קווים ייחודיים וחלקה לרבעים: על גבי הגרפף שורטו שני קווים מקווקווים המציגים את החזיניות של השכר והחזון של הביקוש. הצלבות קווים אלו מחלקת את המפה ארבעה רביעים אסטרטגיים:

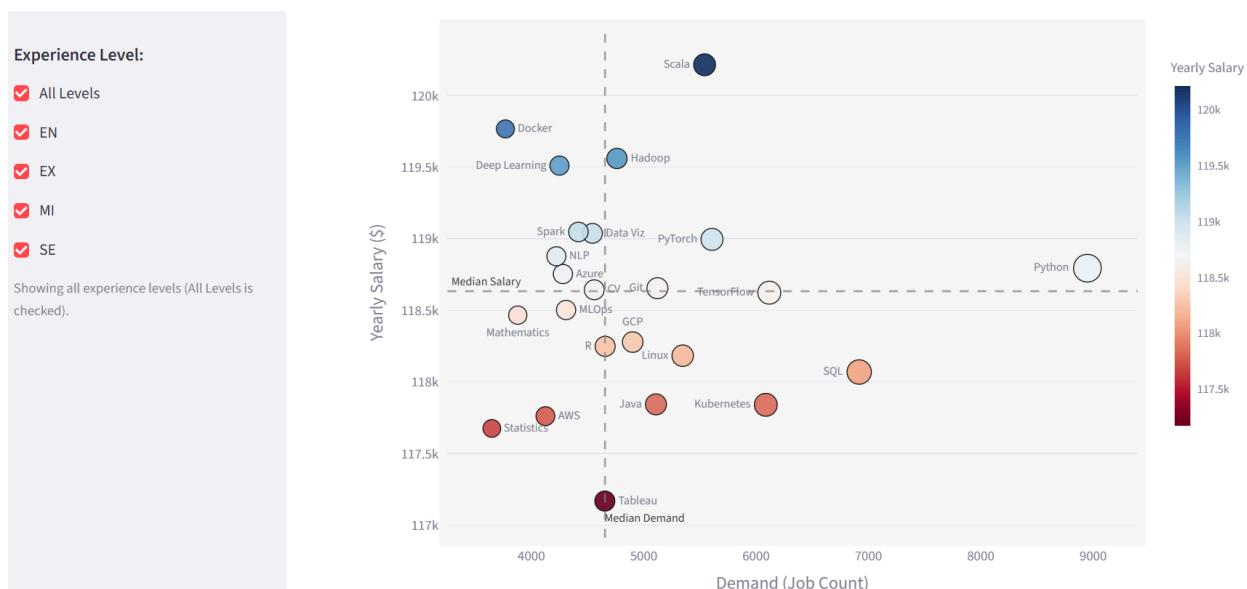
- רביע ימני עליון: מיפויים "כוכבות" - גם שכר גבוה מהממוצע וגם ביקוש גבוה.

- רביע שמאל עליון: מומנויות נישה יקרתיות - שכר גבוה אך ביקוש נמוך יחסית.
- רביע ימני תחתון: מומנויות בסיס - ביקוש גבוה אך שכר נמוך מהמצוע.
- רביע שמאל תחתון: מומנויות בשולטים - שכר וביקוש נמוכים מהמצוע.

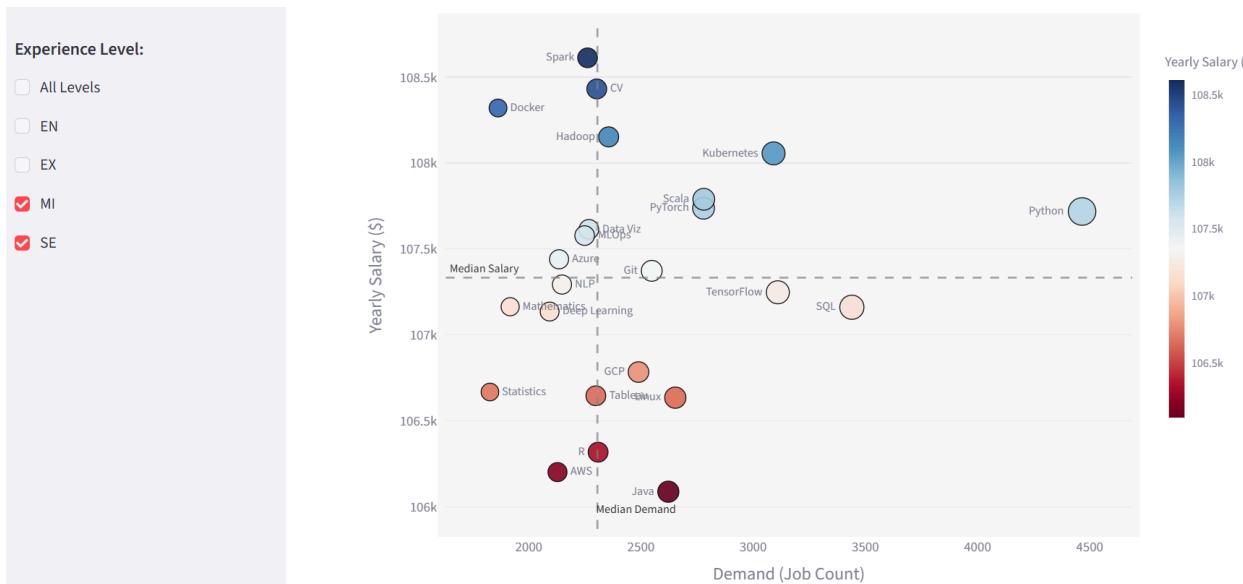
4. אינטראקטיביות ומידע נוספת: הגרף כולל יכולות אינטראקטיביות. במעבר עכבר על כל נקודה, מוצג חלון מידע מפורט היכול את שם המומנות המדוייק, מספר השירותים המדוייק והשכר הממצוע. תוכנה זו מאפשרת לשמר על עיצוב נקי מבליל לאבד את הגישה לנוטונים הגולמיים המדוייקים.



הויזואליזציה נפתחת עם ברירת מחדל של הצגת כל רמות הניסיון.



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



בנוסף, עבור כל קומבינציה של רמות ניסיון מוצגת טבלה המסכם את 5 TOP המומנויות המבוקשות ואת 5 TOP 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 ויזואלייזציה זו נעשה שימוש בספריות הבאות:
  - Streamlit: עבור בניית ממשק המשתמש של אפליקציית-h-Web והציג בחירות המשתמש להציג המדיע המבוקש לפי רמות הניסיון.
  - Pandas: לטובות עיבוד וניהול הנתונים החל מטעינתקובץ הדטה, דרך סינון לפי שנות ניסיון, ועד לביצוע חישובים מורכבים כמו פיצול רשימות הכוורות וחישוב ממוצע שכר ושרותות לפי קבוצות.
  - Plotly Express: עבור יצירת הגרף האינטראקטיבי שמחיש ויזואלית את הקשר בין השכר לביקוש, ומאפשרת למשתמש לרחף מעל הנקודות כדי לראות מידע נוסף, לסנן לרמות הניסיון שמעניינות אותו ולהתמקד בתונונים.

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

### ויזואלייזציה 1: התפלגות שכר לפי תפקיד ורמת ניסיון

א. ביצוע Pre-processing:

בוייזואלייזציה זו לא בוצע עיבוד מוקדם מהותי של הנתונים מעבר לנקיי בסיסי הנדרש לצורך הצגה תקינה. הנתונים נתונים מקובץ CSV המקורי כפי שסופק, ובוצעו רק פעולות מינימליות של המרת שדה השכר (salary\_usd) לערכים מספריים והסרת רישומות חסורת עבר שדות חינויים (שם התפקיד והשכר). לא בוצעו אgregציות, טרנספורמציות, נרמול או סינון מוקדם של הנתונים, על מנת לשמור על נאמנות מלאה להתפלגות המקורית של השכר בשוק העבודה.

ב. שיפור המימוש הויזואלי:

במהלך תהליכי המימוש בוצע שינוי מיוחד הריאשו של הויזואלייזציה. במקור, השוואת השכר בין תפקיד AI התבוססה אך ורק על התפלגות השכר לפי תפקיד. עם התקדמות העבודה ועל בסיס תובנות שעלו מהנתונים, הוחלט להוסיף ממד נוסף לרמת 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 ורמת השכר בכל תפקיד.

בנוסף, נספה אינטראקטיה של סינון לפי סוג העסקה (employment\_type), המאפשרת למשתמש לבחון הבדלים בין משרות Full-time, Part-time, Contract לבין הפכו את הויזואלייזציה מממציע תיאורי בסיסי לכלי אנליטי אינטראקטיבי המאפשר חקירה מעמיקה יותר של הנתונים.

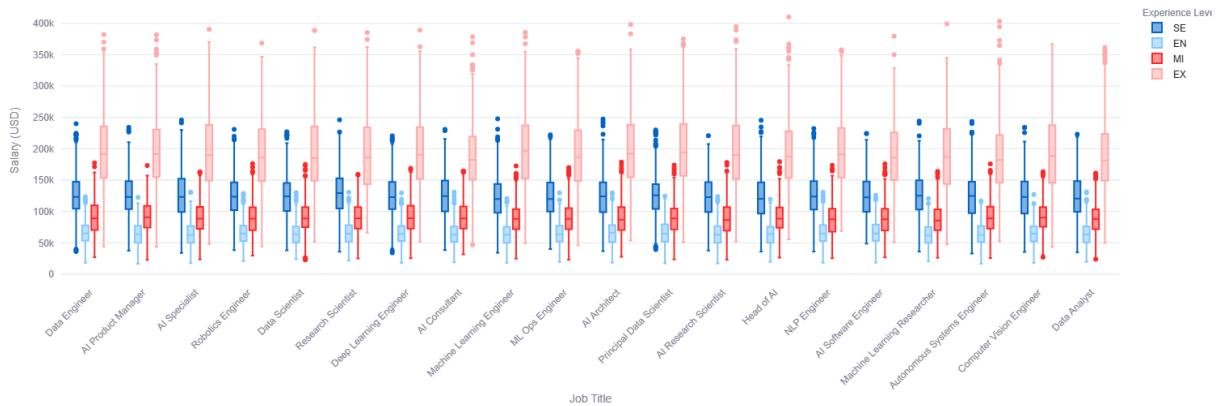
ג. אופן המימוש:

הויזואלייזציה מאפשרת למשתמש אינטראקטיבית 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) המוחושבים בהתאם לנתונים המסוננים, ומאפשרים הינהרעה של השפעת הבחירה שביצע המשתמש.

Salary Distribution (Box Plot)



Filter Data

Select experience\_level

Choose options

Select employment\_type

Choose options

ד. כתיבת הקוד:

הויזואלייזציה מושהה באמצעות שפת Python תוך שימוש בספריות הבאות:

- לבנייתמשק אינטראקטיבי, ניהול סרגלי סינון והציגת התרשיים.
- לטעינת הנתונים, ניקוי בסיסי וסינון דינמי בהתאם לבחירות המשתמש.
- Plotly Express – לייצרתתרשיים Box Plot אינטראקטיביים, כולל קידוד צבע לפי רמת ניסיון והציגת חריגים.

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

- הוספה לוגיקת סינון רב-מדנית
- מיון ציר ה-X לפי חציון שכר
- שילוב מדדי KPI
- התאמת התרשיים לצרכים אנליטיים של חקר שוק העבודה בתחום ה-AI

## ויזואלייזציה 2: Drill-Down לפי מיקום ורמת ניסיון

א. ביצוע Pre-processing:

בווייזואלייזציה זו לא בוצע עיבוד מוקדם מורכב של הנתונים, אלא עיבוד נקודתי לצורכי שימוש ה-Drill-Down. לאחר החלטת המסננים הכלליים (כגון רמת ניסיון וסוג העסקה), בוצע סינון נוסף לפי תפקיד נבחר, ולאחריו חישוב שכר החזינו (Median Salary) לכל צירוף של מדינה ורמת ניסיון. בחירה בשכר החזינו נעודה לצמצם את השפעתם של ערכיים קיצוניים ולאפשר השוואה י-zAmbient;ible בין מדינות. מעבר לכך, לא בוצעו טרנספורמציות או שינוי מבנה נתונים.

ב. שיפור המימוש הויזואלי:

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

בנוסף, הוחלט להוסיף סרגל שליטה המאפשר קביעת דינמיות 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, נסופה אפשרות למין המדינות בציר האנכי לפי רמת ניסיון נבחרת. המין מתבצע על-פי השכר החזינו של אותה רמת ניסיון בכל מדינה, כך שהמדינות מסודרות מהשכר הגבוה לנמוך. כבירות מחדל, המין מתבצע לפי רמת EX(Executive), המייצגת את הדרגה הבכיר ביותר, אך המשמש יכול לבחור למין גם לפי רמות ניסיון אחרות (Mid או Senior, Entry).

ג. אופן המימוש:

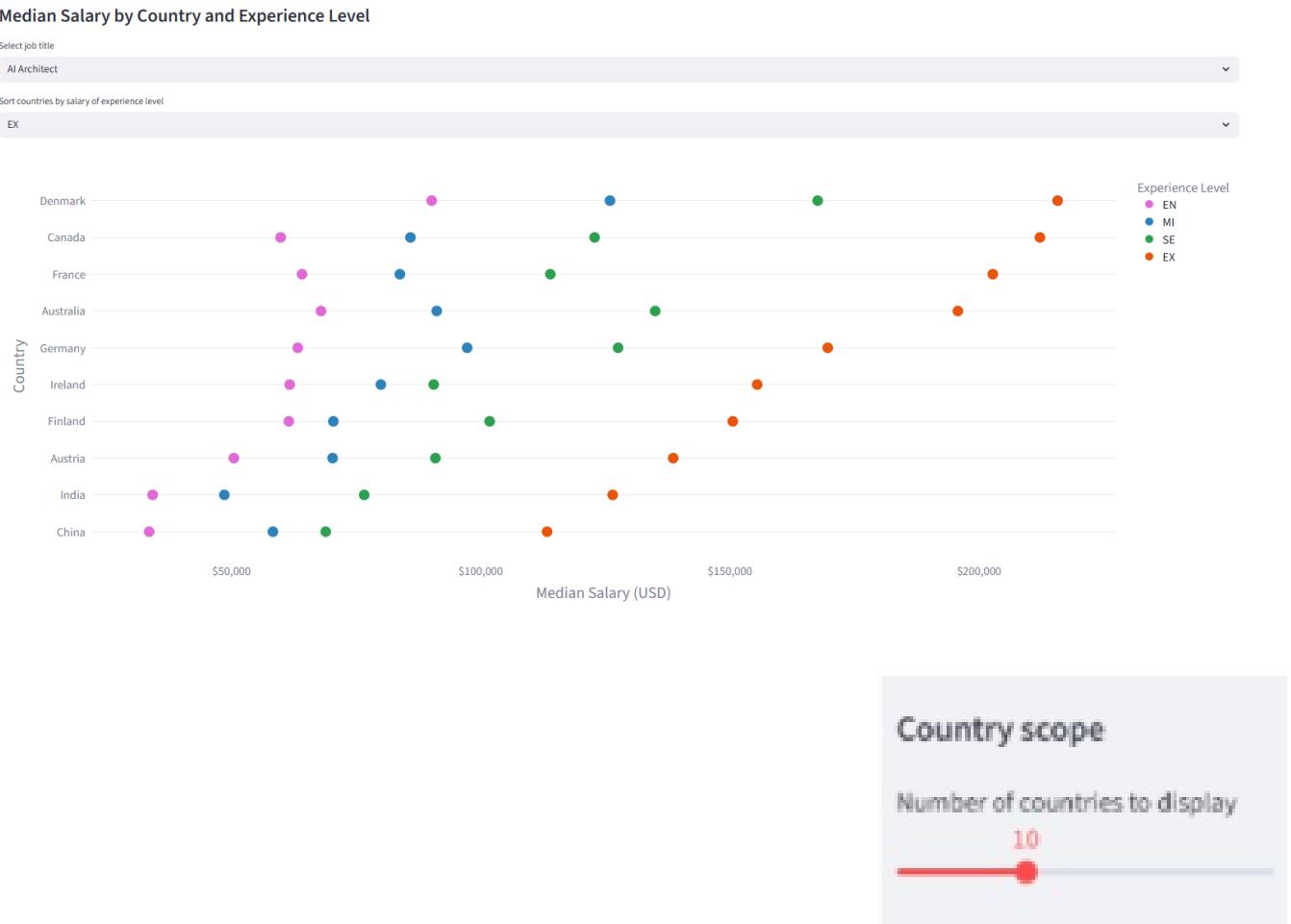
הוויזואלייזציה תומכת במספר שכבות אינטראקטיביה המאפשרות חקירה דינמית של נתוני השכר: בחירת תפקיד באמצעות תיבת בחירה (Select Box).

קביעת מספר המדינות המוצגות באמצעות סרגל הדזה (Slider).

מצגת השכר החזינו לכל מדינה, כאשר הנקודות מקודדות בצבע לפי רמת ניסיון.

בנוסף, נסופה אפשרות למין את המדינות בציר האנכי לפי השכר החזינו של רמת ניסיון נבחרת, כאשר כבירות מחדל המין מתבצע לפי רמת EX..

בעת שינוי אחד מהפרמטרים, התרשים מתעדכן בזמן אמת ומאפשר למשתמש לבדוק כיצד מיקום גיאוגרפי ורמת ניסיון משפיעים יחד על השכר בתפקיד הנבחר.



ד. כתיבת הקוד:

המימוש בוצע באמצעות Python תוך שימוש בספריות:

- Streamlit – לניהול אינטראקטיות משתמש (בחירה תפקיד, סרגל כמות מדינות, פרישת עמודות).
- Pandas – לשינון הנתונים, קיבוץ לפי מדינה ורמת ניסיון וчисוב שכר חיצוני.
- Plotly Express – ליצירת תרשימים פיזור (Scatter Plot) אינטראקטיבי, כולל קידוד צבע לפי רמת ניסיון.

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

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

## ספריות חיצונית ושימוש בAI:

עבור העבודה נעשה שימוש בספריות החיצונית:

- Streamlit - ניהול אינטראקטיביות משתמש
- Pandas - לטין הנ נתונים
- Plotly Express - לייצור התרשיים

כל ויזואלייזציה עשויה שימוש בכלים אלו באופן המותאם לה.

נעשה שימוש בAI באופן הבא: תחילת חשבנו מה נרצה שהויזואלייזציה תציג במדוייק וכייד תראה. לאחר מכן שלחנו לAI את הדרישות המדוייקות לייצור הויזואלייזציה - לרבות הכלים בו נעבד, ממה מורכבת הויזואלייזציה במדוייק, סקלאלת צבעים, האופן האיטרטיבי בו יעבד וכו' - ככלmor תיאור מكيف של כל מה שהיא בויזואלייזציה כך שהAI יממש את הקוד שמבצע את כל הדרישות של הויזואלייזציה.