**שלום מר וודבריג,אני ושני השותפים לפרוייקט במדעי הנתונים רצינו לוודא איתך אם אפשר להתחיל לעבוד על הפרוייקט בנושא הבא:**

**נושא: חיזוי תמ"ג + נתונים על המדינה**

**שאלה ראשית:** האם ניתן לחזות את שינוי התמ"ג בהתחשב בגורמים שונים (מס' תושבים, אבטלה, דירוג איכות החינוך, אוצרות לאומיים, ייצוא הייטק, מצב ביטחוני, איכות הרפואה, מחירי הדירות/שכירות, איכות החיים ומחירי מקדונלדס)  
  
**שאלה שנייה:** האם ניתן לחזות את קצב גידול האוכלוסייה בהינתן דירוגי: איכות חיים, חינוך, רפואה ומצב בטחוני?

**שאלה שלישית:**

לחלק את המדינות לקבוצות בעזרת למידת מכונה לא מונחית (clustering)

שאלה ראשית

האם ניתן לחזות את שינוי התמ"ג בהתחשב [מס' תושבים](https://data.worldbank.org/indicator/SP.POP.TOTL?locations=1W),[אבטלה](https://www.theglobaleconomy.com/rankings/Unemployment_rate/) ,[דירוג איכות החינוך](https://worldpopulationreview.com/country-rankings/education-rankings-by-country), [אוצרות לאומיים](https://www.theglobaleconomy.com/rankings/coal_reserves/), [ייצוא הייטק](https://data.worldbank.org/indicator/TX.VAL.TECH.MF.ZS), [מצב ביטחוני](https://en.wikipedia.org/wiki/List_of_ongoing_armed_conflicts), [איכות הרפואה](https://www.numbeo.com/health-care/rankings_by_country.jsp), [מחיר דירות/שכירות](https://www.numbeo.com/property-investment/rankings_by_country.jsp), [מחירי מקדונלדס](https://www.numbeo.com/cost-of-living/country_price_rankings?itemId=3), [איכות חיים](https://www.numbeo.com/quality-of-life/rankings_by_country.jsp?title=2021)?

שאלת משנית

האם ניתן לחזות את קצב גידול האוכלוסייה בהינתן [דירוג איכות החינוך](https://worldpopulationreview.com/country-rankings/education-rankings-by-country), [מצב ביטחוני](https://en.wikipedia.org/wiki/List_of_ongoing_armed_conflicts), [איכות הרפואה](https://www.numbeo.com/health-care/rankings_by_country.jsp), [איכות חיים](https://www.numbeo.com/quality-of-life/rankings_by_country.jsp?title=2021)?

ניתוח נתונים:

1. קורלציה בין הנתונים:

* איכות חיים לילודה.
* רפואה ותמותה.
* בין תמ"ג לכולם כדי לראות האם יש קשר.
* קשר בין חינוך להייטק.

1. גרף פאי של מדינות לפי תמ"ג
2. גרף עמודות של תמ"ג לפי יבשות
3. גידול האוכלוסייה לפי יבשות

התמודדות עם נתונים חסרים:

* לקחת X מדינות עם הכי הרבה מידע
* להתעלם ממידע חסר בנושאים: "מחירי מקדונלדס"

מניפולציות על המידע בהמשך:

* להשתמש במדינות שנמצאות באותה קבוצת חלוקה שהמחשב חילק בלמידה לא מונחית.
  + בדיקה מדוע המדינות חולקו בצורה כזו

**שאלות אופציונליות:**

* האם ניתן לחזות השמנת יתר בהינתן המידע הנ"ל על המדינה?
* האם ניתן לחזות את מחיר המקדונלדס במדינה?

**TODO:**

1. לחלק ולאסוף את כל המאגרים לרשימה עם כותרות (מאיפה לוקחים).
2. לסדר + להבין על מה לעשות סקרייפינג בכל מקור.
3. בניית DBים.
4. לפתוחhttps://github.com/tonyhasson/ETZ-GDP
5. ניקוי נתונים + חריגים ושקרנים.
6. הצגה ויזואלית + קורולציה : אנליזה על הנתונים + ניתוח חקרני ומעמיק.
7. חלוקה לקלאסטרים ע"י לימדה לא מונחית.
8. עבור חיזוי נשתמש בלמידה לא מונחית

מקורות:

<https://www.guru99.com/mongodb-vs-mysql.html>

יש פה משהו שמגיע לנו על מונגו:

<https://education.github.com/pack>