תרגיל 4

נא להגיש את התרגיל בזוגות

יש להגיש בקובץ WORD

עבור כל תרגיל יש להגיש את הקוד בפייטון + צילומי מסך של התוצאות .

שימו לב- תרגילים ללא צילומי מסך לא יבדקו!

האחראית על התרגיל :אביגיל שטקל [Avigail.st@gmail.com](mailto:Avigail.st@gmail.com)

שאלה 1 – SPARK

מצורף קובץ JSON עם רשימת ספרים בספריה

את התרגיל הזה יש לכתוב אך ורק עם פונקציות של SPARK בלבד (לסעיף ב' אפשר להשתמש בFOREACH אך ורק כדי להדפיס את התוצאה)

1. הדפס טבלה עם כל הספרים של סופרים שמתחילים בF. הנתונים שיש להדפיס : שם הספר , שם הסופר ומספר השנים שעברו מאז שהוא יצא לאור.
2. יש לחשב את מספר הדפים הממוצע שכתב כל סופר אבל רק ספרים באנגלית

שאלה 2 – LINEAR REGRESSION

יש להשתמש בפייטון ולממש את ה-classifiers מ - scratch

לא להשתמש בשום ספריה של learning machine

כמובן שאפשר להשתמש ב-numpy ולהתבסס על השקפים בשיעור

מצ"ב רשימת נתונים על מחירי דירות בעיר אחת בחו"ל:  
הנתונים הם:

1מחירי המכירה המקומיים, במאות דולרים;  
2, מספר חדרי האמבטיה;  
3, שטח האתר באלפי מטרים רבועים;  
4 גודלו של שטח המחיה באלפי מטרים רבועים;  
5 מספר המוסכים;  
6, מספר החדרים;  
7, מספר חדרי שינה;  
8, הגיל בשנים;  
9, סוג הבנייה  
10, סוג אדריכלות  
11, האם יש מכבי אש (0 אין 1 יש).  
12.מחיר המכירה

קראו את הנתונים ב- python ו:

1. בנו מודל **linear** regression בפייטון אשר חוזה את מחיר הדירה על הנתונים 1-11 .

השתמשו ב-75% מהנתונים עבור בניית המודל, ב-25% הנותרים בידקו את המודל שיצרתם .

1. דווחו את התוצאות ב-**mean squared-error .** יש להשוות את התוצאות של המודל לנתונים הקיימים .
2. בנו **logistic** regression classifier בפייטון אשר חוזה אם יש מקומות לכיבוי אש (נתון מס11). השתמשו ב-%85 מהנתונים עבור בניית המודל, ב-15% הנותרים בידקו את המודל שיצרתם .

4. דווחו את התוצאות במונחים של **accuracy, recall, precision and F-measure**