**פרויקט בינה עסקית – הלוואות**

**קורס:**

בינה עסקית – מס' קורס 234020500

שנת ה'תשפ"ג – סמסטר ב'

**שמות הסטודנטים בפרויקט:**

עידן לוי – 208995191  
ליאור ברשדסקיי – 319000337

**מרצה מוביל הפרויקט:**אור יצחק פרץ

**חלק 1: הבנה עסקית**

אוסף הנתונים נלקח מאתר Kaggle.

הנתונים עוסקים באישור למתן הלוואה על בסיס הכנסה שנתית של בני זוג.

<https://www.kaggle.com/datasets/mirzahasnine/loan-data-set>

**חלק 2: הבנת הנתונים**

בהתבסס על נתונים אלו, הועלו 3 שאלות עסקיות אותן נרצה להעלות מניתוח הנתונים:

1. מהו סוג האזור שבו סך ההלוואות שנלקחו הוא הגבוה ביותר וכמה כמות זו מהווה באחוזים מסך ההלוואות שנלקחו?
2. האם קיים קשר ישר/קשר הפוך/אין קשר בין מספר ילדים לגובה ההלוואה ומה עוצמתו של קשר זה חזק/בינוני/חלש?
3. מה הרבעון שבו ממוצע ההלוואות הוא הגבוה ביותר אצל נשואים?

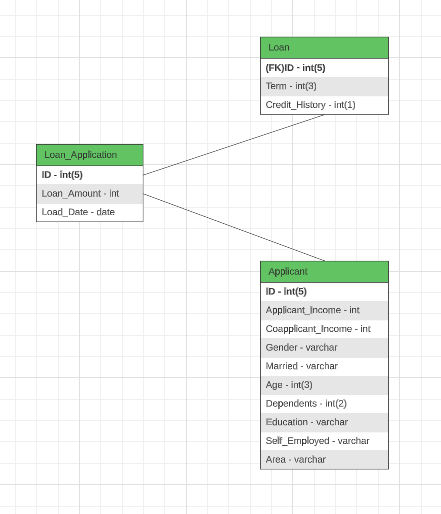
לטובת המענה על שאלות אלו, נבחן מספר KPI's שיהוו את המדדים עבור שאלות אלו:

1. סכום כולל של ההלוואות שנלקחו.
2. סכום הלוואות בחלוקה לסוג-אזור.
3. מספר הלווים הכולל.
4. מספר הלווים בחלוקה לסוג-אזור.
5. ממוצע הלוואות בחלוקה לרבעונים.

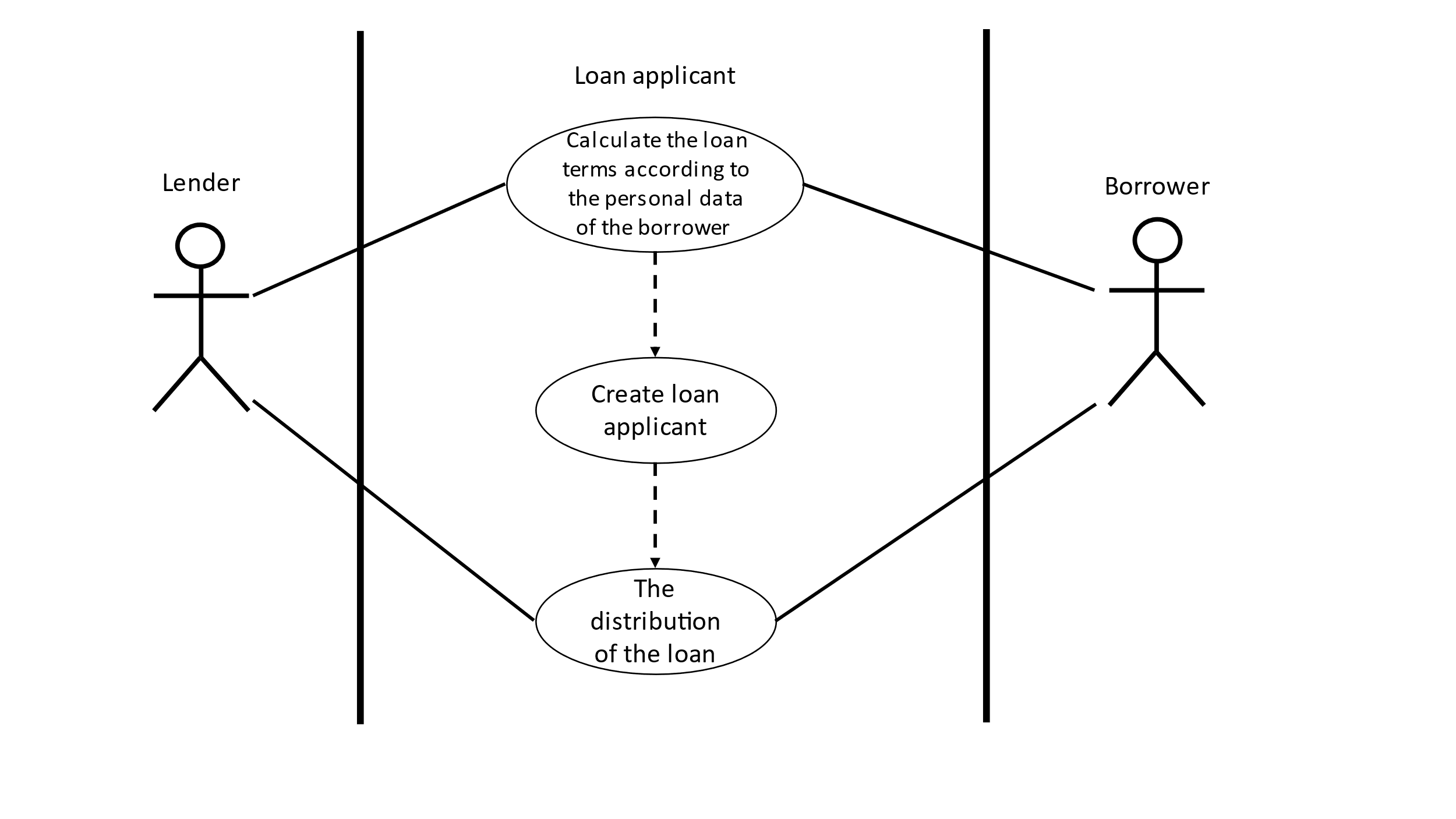
**חלק 3: הכנת הנתונים – ניקוי וסידור הנתונים**

תחילה קראנו את הנתונים לקובץ פייתון ובדקנו את כמות השורות והעמודות בקובץ. לאחר מכן ניקינו מן הנתונים את השורות בהן קיימים תאים שחסר בהם מידע.  
מצאנו כי כל העמודות רלוונטיות לנו ולכן בחרנו שלא להוריד אף עמודה ממס"ד הנתונים שברשותנו.  
בנוסף, הורדנו כפילויות בנתונים ושורות בהן קיים מידע של טקסט במקום שציפינו לקבל ממנו מספר ולהפך.  
לבסוף יצרנו תרשים שמציג את התפלגות סכומי ההלוואות שניתנו וחישבנו טווח בין רבעוני עבור עמודה זו.

**תרשים ERD**



**תרשים use-case**



לדעתנו מס"ד הנתונים יחסית עשיר ומאפשר תמיכה בתהליך קבלת ההחלטה על גובה ההלוואה הצפוי בהתבסס על הנתונים שברשותנו. התהליך דורש קבלת מידע ראשוני מהלווה, המידע מנותח על פי הנתונים שברשותנו ולבסוף מתקבל סכום הלוואה שתואם לנתוני הלווה. התהליך לא אמור לקחת יותר ממספר דקות בודדות על מנת לקבל הצעה לקבלת הלוואה.

**חלק 4: מודל**

הסכמה שנבחרה היא סכמת Star. סכמה זו מתאימה למס"ד הנתונים שלנו מכיוון שסכמת הנתונים בנויה באופן של טבלה מרכזת אחת שממנה יוצאות מספר טבלאות משנה שאינן קשורות באופן היררכי כלשהו. למשל עבור מספר מזהה של כל הלוואה שניתנה, ניתן לראות את הגיל של לוקח ההלוואה, את ההכנסה של בעל ההלוואה, את החודשים להלוואה וכו'.

**חלק 5: הערכה**

**חלק 6: יישום**

**נפסח א' – מס"ד הנתונים**