

# פרויקט ניתוח סיכון אשראי

הפרויקט מבוסס על [מערך נתונים סיכון אשראי באתר kaggle.com](https://www.kaggle.com/c/national-health-survey-israel) המציגים היסטוריות הלואות של לקוחות. הנתונים מאפשרים לבצע ניתוח סיכון אשראי ולבחן אילו לקוחות נוטים יותר להגיע לחדרות פירעון, כמו גם לחזות יכולת ההחזר של לקוחות בעtid. לבתו הגוננים והבנת סיכון האשראי, ניסיתי לזהות דפוסים ומאפיינים המבדילים לוויים מסוכנים מלוויים יציבים, ומכאן עלן שאלת המחקר ושאלות המשנה.

## שאלת המחקר: מה הפרופיל של לווה מסוכן?

שאלות שנעזרתי בהם כדי לענות על שאלת זו:

- האם לוויים עם היסטורית חדרות פירעון מצליכים להחזיר הלואות חדשות?
- האם לוויים צעירים מסוכנים יותר מלוויים מבוגרים?
- האם יש קשר בין החזר מההכנסה לוותק תעסוקתי?
- האם סכום גבוה מגדיל את סיכון ההלוואה?
- האם דירוג ההלוואה מנבأ את סיכון ההחזר?

כדי להבין את סיכון האשראי בצורה מקיפה, התחלתי בניתוח ריאוני באקסל להסרת ערכים חריגים ולא הגיוניים. לאחר מכן, המשכתי בארגון הנתונים בפאננד, עברתי ל-MySQL לביצוע שאלות מורכבות וניתוח דפוסים בין טבלאות, ולבסוף השתמשתי בטאבלו להצגת התובנות בצורה ייזואלית ונוחה להבנה.

## פאנדו:

הכנת נתונים:

טיפולתי בנתוני האשראי על ידי מילוי הערכים החסרים בריבית הלוואה לפי הממוצע של כל דירוג הלוואה, ובערכים החסרים של ותק העבודה רשותי 0. בנוסף, יצרתי עמודת יחס חוב להכנסה והפקתי קודים למלילים ברורות ( Y/N הפכו ל-No/Yes ו-0/1 הפכו ל- Non-Default .(/ Default

```
import pandas as pd

#טעינה הנתונים
df = pd.read_csv('credit_risk_dataset.csv')

#מילוי ריבית חסраה לפי ממוצע של כל דירוג
df['loan_int_rate'] =
df.groupby('loan_grade')['loan_int_rate'].transform(lambda x:
x.fillna(x.mean()))

#מילוי ותק עבודה 0 בערכים חסרים
df['person_emp_length'] = df['person_emp_length'].fillna(0)

#החלפת 0 ו-1 למלילים ברורות (Default / Non-Default)
df['loan_status'] = df['loan_status'].replace({0: 'Non-Default', 1:
'Default'})

#颠倒 N-ו Y-ו החלפת
df['cb_person_default_on_file'] =
df['cb_person_default_on_file'].replace({'N': 'No', 'Y': 'Yes'})

#יצירת עמודת יחס חוב להכנסה
df['debt_to_income'] = df['loan_amnt'] / df['person_income']

#שמירת הקובץ
df.to_csv('cleaned_credit_data.csv', index=False)
```

פיזול הקובץ:

יצרתי מזהה ייחודי לכל לוהה, בניתי מילון סוג הלוואות שמאפשר שינוי כל של סוג ההלוואה מבלי לעורר ידנית, ופיצלתי את הנתונים לשולשה גילוונות: פרטי הלקוח, פרטי ההלוואה ומילון סוג הלוואות.

```
import pandas as pd
import os

desktop_path = os.path.join(os.path.expanduser("~"), "Desktop",
"Credit_Project_Final_15.xlsx")

#טעינה הנתונים
df = pd.read_csv('credit_risk_data.csv')
df['person_id'] = range(1, len(df) + 1)

#יצירת המילון
loan_types = pd.DataFrame({
    'type_id': range(1, len(df['loan_intent'].unique()) + 1),
    'loan_intent': df['loan_intent'].unique()
})
df['type_id'] =
df['loan_intent'].map(dict(zip(loan_types.loan_intent,
loan_types.type_id)))

#פיזול לגילוונות:
persons = df[['person_id', 'person_age', 'person_income',
'person_home_ownership',
'person_emp_length', 'cb_person_cred_hist_length']]

loans = df[['person_id', 'type_id', 'loan_grade', 'loan_amnt',
'loan_int_rate',
'loan_status', 'loan_percent_income',
'cb_person_default_on_file', 'debt_to_income']]
```

#שמירת הקובץ

```

with pd.ExcelWriter(desktop_path) as writer:
    persons.to_excel(writer, sheet_name='Persons', index=False)
    loans.to_excel(writer, sheet_name='Loans', index=False)
    loan_types.to_excel(writer, sheet_name='Loan_Types',
index=False)

```

## טבלאות:

**טבלה Persons – לקוחות**

תיאור בעברית	שם عمودה
מזהה לקוחות	person_id
גיל	person_age
הכנסה שנתית	person_income
בעלות על דירה	person_home_ownership
ותק תעסוקתי (בשנים)	person_emp_length
אורך היסטורית אשראי (בשנים)	cb_person_cred_hist_length

  

**טבלה Loans – הלוואות**

תיאור בעברית	שם عمودה
מזהה לקוחות	person_id
מזהה סוג הלוואה	type_id
דרוג הלוואה	loan_grade
סכום הלוואה	loan_amnt
ריבית הלוואה	loan_int_rate
סטטוס הלוואה	loan_status
אחוז החזר מהכנסה	loan_percent_income
היסטוריה חקלות פירעון	cb_person_default_on_file
יחס חוב להכנסה (DTI)	debt_to_income

  

**טבלה Loan\_Types – סוג לקוחות**

תיאור בעברית	שם عمودה
מזהה סוג הלוואה	type_id
מטרת הלוואה	loan_intent

## :SQL

- קודם כל בדקי השדות מלאים, כדי לוודא שאין ערכים חסרים שuvwלפוגע בניתוח הנתונים.

```
select *
from persons
where person_id is null
or person_age is null
or person_income is null
or person_home_ownership is null
or person_emp_length is null
or cb_person_cred_hist_length is null

select *
from loans
where person_id is null
or type_id is null
or loan_grade is null
or loan_amnt is null
or loan_int_rate is null
or loan_status is null
or loan_percent_income is null
or cb_person_default_on_file is null
or debt_to_income is null

select *
from loan_types
where type_id is null
or loan_intent is null
```

- חילקתי את הלוויים לקבוצות גיל לפי עשרים ובחנותי עברו כל קבוצה את כמות הלוויים, ההכנסה החזינית, יחס חוב הכנסה, אחוז השוכרים ושיעור חדלות הפירעון, כדי לבדוק האם לוויים צעירים מסוימים יותר ממוגרים.

```

select
case
when p.person_age >= 60 then '60+'
else concat(floor(p.person_age / 10) * 10, '-',
floor(p.person_age / 10) * 10 + 9)
end as age_group_text,
count(*) as borrower_count,
percentile_cont(0.5) within group (order by p.person_income) as
median_income,
avg(l.debt_to_income) as avg_dti,
avg(case when p.person_home_ownership = 'RENT' then 1.0 else 0.0
end) as renter_percentage,
avg(case when l.loan_status = 'Default' then 1.0 else 0.0 end)
as default_rate
from persons p
join loans l on p.person_id = l.person_id
group by
case
when p.person_age >= 60 then '60+'
else concat(floor(p.person_age / 10) * 10, '-',
floor(p.person_age / 10) * 10 + 9)
end
order by
min(p.person_age)

```

- חילקתי הלוואות לפי מטרת ההלוואה ובחנותי עברו כל מטרה את סכום ההלוואה הממוצע, שיעור חדלות הפירעון ומספר הלוואות, כדי לבדוק האם סכומי הלוואה גבוהים יותר מגדילים את סיכון ההלוואה.

```

select
t.loan_intent,
round(avg(l.loan_amnt), 0) as avg_loan_amount,

```

```

round(avg(case when l.loan_status = 'Default' then 1 else 0 end)
* 100, 2) || '%' as default_rate,
count(l.person_id) as total_loans
from loans l
join loan_types t on l.type_id = t.type_id
group by t.loan_intent
order by avg_loan_amount desc

```

- חילקתי את הלוואות לפי דירוג ובחןתי עבור כל דירוג את מספר הלוואות ושיעור חדלות הפירעון, כדי לבדוק האם דירוג הלוואה מניבא את סיכון ההחזר.

```

select
loan_grade,
count(person_id) as total_loans,
round(avg(case when loan_status = 'Default' then 1 else 0 end) *
100, 2) as default_rate
from loans
group by loan_grade
order by loan_grade asc

```

## טבלו:

בטאבלו יצרתי גרפים שמחמישים את שאלות המחבר ומוסיעים להבין את הפרופיל של לקוחות הנוטים יותר לחדלות פירעון.

## **מסקנות הניתוח:**

מאפייני לווה בסיכון גבוה: אדם שהגיע בעבר לחדלות פירעון, בגילאי 29-40 או מעל גיל 60, שלווה הלוואה לאיחוד חובות, בעל ותק עבודה נמוך מ-4 שנים ודירוג הלוואה שלו סומת.