Loan Delinquencies in Analysis

Understanding the roles of credit, inflation, and interest rates on loan delinquency

Client: Federal Student Aid

Business Analysis (yz2433, xh393,izd3

Table of contents

```
import pandas as pd
import numpy as np
```

Importing the credit data Columns we are keeping:

- userID: unique respondent in-survey identification number used to match respondents between different surveys

- date: data survey was conducted (2013 - 2024)

- N1_3: whether they possess student loan or not (1 for possess)

- N1_4: whether they possess home-based loan or not (1 for possess)

- N1_5: whether they possess auto-based loan or not (1 for possess)

- N1_6: whether they possess other personal loan or not (1 for possess)

- N2_1: current balance of credit card loans

- N2_3: current balance of student loans

- N10_7: Dollar amount granted by lender for student loan

- N15: Loan payments late by more than 30 days in past 12 months (1 for late payment)

- N16: Loan payments late by more than 90 days in past 12 months (1 for late payment)

- N17a_5: How likely will request credit card limit increase in next 12 months

- N22: Credit score (1 for below 620; 2 for 620 - 679; 3 for 680 - 719; 4 for 720 - 760; 5 for Above 760; 6 for don't know)

- N23: Last time checked credit score (1 for <1 mon; 2 for between 1 - 6 mon ago; 3 for 6-12 mon ago; 4 for 1-2 yrs ago; 5 for >2yrs ago; 6 for don't know)

- N25: Likelihood of being able to come up with $2000 in the next month given an unexpected need (0-100 percent)

```python
credit_df = pd.read_excel("potentialDataSetsHW2/sce_credit.xlsx",
sheet_name="Data")
reduced_credit_df = credit_df[["userid", "date", "weight",
                               "N1_3", "N1_4","N1_5","N1_6",
                               "N2_1", "N2_3",
                               "N10_7", "N15", "N16", "N17a_5",
                               "N22", "N23", "N25"]]
reduced_credit_df
```

	userid	date	weight	N1_3	N1_4	N1_5	N1_6	N2_1	N2_3	N10_7	N15	N16	N17a_5	N22	N23	N25
0	7000030	7310	5.3820	100	0	0	0	NaN	3600.0	NaN	1.0	NaN	NaN	NaN	NaN	NaN
1	7000034	7310	0.5570	100	0	1	0	200000	100000	NaN	0.0	NaN	NaN	NaN	NaN	NaN
2	7000320	7310	0.8680	00	0	0	0	NaN	NaN	NaN	0.0	NaN	NaN	NaN	NaN	NaN
3	7000320	7310	0.4220	00	0	0	0	70000	NaN	NaN	0.0	NaN	NaN	NaN	NaN	NaN
4	7000320	7310	0.6380	00	0	0	0	NaN	NaN	NaN	0.0	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3435775	017362400		0.7417	07	1	1	0	250.0	50000	NaN	0.0	NaN	2.0	4.0	4.0	93.0
3435875	017380400		2.2620	060	0	0	0	8000.0	NaN	NaN	0.0	NaN	NaN	5.0	1.0	100.0
3435975	017384400		2.4673	106	0	1	0	100001	8000	NaN	0.0	NaN	NaN	1.0	1.0	0.0
3436075	017388400		0.5450	103	0	0	0	190005	2000	NaN	0.0	NaN	NaN	2.0	5.0	100.0
3436175	017803400		0.3823	179	0	1	0	NaN	95000	NaN	0.0	NaN	NaN	5.0	1.0	99.0

```python
reduced_credit_df.columns
```

```
Index(['userid', 'date', 'weight', 'N1_3', 'N1_4', 'N1_5', 'N1_6', 'N2_1',
       'N2_3', 'N10_7', 'N15', 'N16', 'N17a_5', 'N22', 'N23', 'N25'],
      dtype='object')
```

```python
#reduced_credit_df["year"] = reduced_credit_df["date"].astype(str).str[:4]
```

Importing the housing spending data Columns we are keeping:

- userID: unique respondent in-survey identification number used to match respondents between different surveys

- date: data survey was conducted (2013 - 2024)

- qsp3_1: home appliance purchase during last 4 months (only have data from 2015 and on; house option added only Aug 2016)

- qsp3_2: electronics, computers, or cell phone purchase during last 4 months (only have data from 2015 and on; house option added only Aug 2016)

- qsp3_3: furniture purchase during last 4 months (only have data from 2015 and on; house option added only Aug 2016)

- qsp3_4: home repairs, improvements or renovations purchase during last 4 months (only have data from 2015 and on; house option added only Aug 2016)

- qsp3_5: cars/vehicle purchase during last 4 months (only have data from 2015 and on; house option added only Aug 2016)

- qsp3_6: trip/vacation purchase during last 4 months (only have data from 2015 and on; house option added only Aug 2016)

- qsp3_9: house/apartment purchase during last 4 months (only have data from 2015 and on; house option added only Aug 2016)

- qsp3_8: no large purchase during last 4 months (only have data from 2015 and on; house option added only Aug 2016)

- qsp5_1: proportion of current monthly salary on housing

- qsp5_2: proportion of current monthly salary on utilities

- qsp5_3: proportion of current monthly salary on food

- qsp5_4: proportion of current monthly salary on clothing, footwear, personal care

- qsp5_6: proportion of current monthly salary on transportion

- qsp5_7: proportion of current monthly salary on medical care

- qsp5_8: proportion of current monthly salary on education and child care

- qsp5_9: proportion of current monthly salary on other (gifts, child support, charity, etc.)

- qsp12_n: if get 10% more income, what they would do (3 for use all to pay debt; 5,6, 7 for use some to pay debt; 1, 2, 4 for use none for debt)

- qsp15new: variability in household income month to month (1 for <5%; 2 for 5%-15%; 3 for >15%)

- k2e: whether there exist a family budget (1 for yes)

```
""" hhspending_df = pd.read_excel("potentialDataSetsHW2/
sce_household_spending.xlsx", sheet_name="Data")
reduced_hhspending_df = hhspending_df[["userid", "date",
                                    "qsp3_1", "qsp3_2", "qsp3_3",
"qsp3_4","qsp3_5","qsp3_6","qsp3_9","qsp3_8",
                                    "qsp5_1",
```

```
"qsp5_2","qsp5_3","qsp5_4","qsp5_5", "qsp5_6", "qsp5_7", "qsp5_8", "qsp5_9",
                                  "qsp12n", "qsp15new", "k2e"]]
reduced_hhspending_df """
```

```
' hhspending_df = pd.read_excel("potentialDataSetsHW2/
sce_household_spending.xlsx", sheet_name="Data")\nreduced_hhspending_df =
hhspending_df[["userid", "date", \n
"qsp3_1", "qsp3_2", "qsp3_3", "qsp3_4","qsp3_5","qsp3_6","qsp3_9","qsp3_8",\n
"qsp5_1", "qsp5_2","qsp5_3","qsp5_4","qsp5_5", "qsp5_6", "qsp5_7", "qsp5_8",
"qsp5_9",\n                                    "qsp12n", "qsp15new",
"k2e"]]\nreduced_hhspending_df '
```

```
#reduced_hhspending_df["year"] =
reduced_hhspending_df["date"].astype(str).str[:4]
```

Importing Job and Income Data

Columns we want to keep:

- userid: Unique identifier

- date: date survey was conducted (2014 and on)

- l3: annual salary before tax and deductions (including bonuses, overtime, tips/commissions)

- l7dk_1: never had a paying job (1 for never; empty for having a paying job)

```
""" job_df = pd.read_excel("potentialDataSetsHW2/sce_labor.xlsx",
sheet_name="Data")
reduced_labor_df = job_df[["userid", "date",
                                  "l3", "l7dk_1"]]
reduced_labor_df """
```

```
' job_df = pd.read_excel("potentialDataSetsHW2/sce_labor.xlsx",
sheet_name="Data")\nreduced_labor_df = job_df[["userid", "date", \n
"l3", "l7dk_1"]]\nreduced_labor_df '
```

```
#reduced_labor_df["year"] = reduced_labor_df["date"].astype(str).str[:4]
```

Importing household status

Columns we want to keep:

- userid

- date

- a1new: current living situation (1 for married; 2 for separated; 3 for divorced; 4 for widowed; 5 for never married)

- b1_1: household income

- b5_1: Percentage change in HH income last 12 mos. vs. previous year

- i1_6: Disabled

```
""" hhfinance_df = pd.read_excel("potentialDataSetsHW2/
sce_household_finance.xlsx", sheet_name="Data")
reduced_hhfinance_df = hhfinance_df[["userid", "date",
                                    "a1new", "b1_1", "b5_1", "i1_6"]]
reduced_hhfinance_df """
```

```
' hhfinance_df = pd.read_excel("potentialDataSetsHW2/
sce_household_finance.xlsx", sheet_name="Data")\nreduced_hhfinance_df =
hhfinance_df[["userid", "date", \n
"a1new", "b1_1", "b5_1", "i1_6"]]\nreduced_hhfinance_df '
```

```
#reduced_hhfinance_df["year"] =
reduced_hhfinance_df["date"].astype(str).str[:4]
```

```
""" reduced_credit_df.set_index(["userid", "date"], inplace = True)
reduced_hhspending_df.set_index(["userid", "date"], inplace = True)
reduced_labor_df.set_index(["userid", "date"], inplace = True)
reduced_hhfinance_df.set_index(["userid", "date"], inplace = True) """
```

```
' reduced_credit_df.set_index(["userid", "date"], inplace =
True)\nreduced_hhspending_df.set_index(["userid", "date"], inplace =
True)\nreduced_labor_df.set_index(["userid", "date"], inplace =
True)\nreduced_hhfinance_df.set_index(["userid", "date"], inplace = True) '
```

```
""" joined_df = reduced_credit_df.join(reduced_hhspending_df, how = "inner")
"""
```

```
' joined_df = reduced_credit_df.join(reduced_hhspending_df, how = "inner") '
```

```
""" common_index = set(reduced_credit_df.index) & \
             set(reduced_hhfinance_df.index)

print(f"Number of common (userid, date): {len(common_index)}") """
```

```
' common_index = set(reduced_credit_df.index) &
set(reduced_hhfinance_df.index) \n\nprint(f"Number of common (userid, date):
{len(common_index)}") '
```