IST 707: Applied Machine Learning

Name: Chaithra Kopparam Cheluvaiah

SUID: 326926205

Email: ckoppara@syr.edu

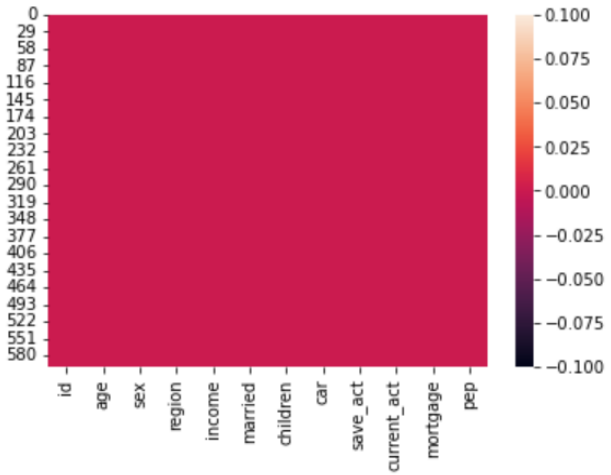
AML Homework 03 – Association Rule Mining

DATA

Banking data is collected by marketing department of a financial firm. It includes customer demographic information and type of accounts. Data is in **CSV** format. Data is loaded to python application as a data frame. It has **600 rows** and **12 columns**. Data is very much clean – no missing values, null values, or NAs found in the dataset.

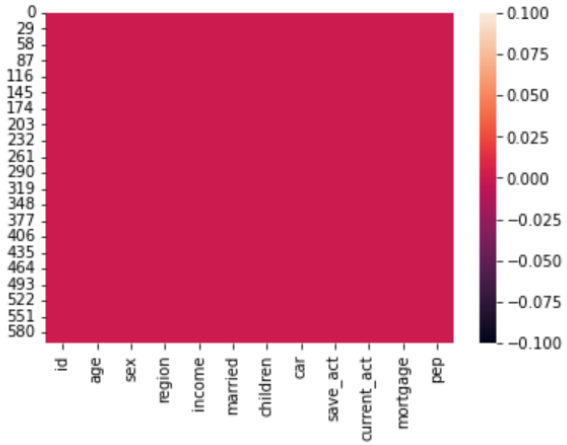
```
sns.heatmap(bank.isnull())
```

```
<AxesSubplot:>
```




```
sns.heatmap(bank.isna())
```

```
<AxesSubplot:>
```



IST 707: Applied Machine Learning

Name: Chaithra Kopparam Cheluvaiah

SUID: 326926205

Email: ckoppara@syr.edu

DATA PRE-PROCESSING REQUIRED FOR RULE MINING

1. **Removing the 'id' column:** Id column is used in uniquely identifying each record in the dataset. We cannot find any meaningful pattern with this attribute hence it has been removed.

```
# drop id column
bank.drop(['id'], axis=1, inplace=True)
```

2. **Discretization**: Converted numeric variables – income, age to categorical variables by binning. Also, renamed the values in the columns – married, children, car, savings account, current account, mortgage, and PEP to have meaningful representation in the rules.

```
# discretization: converting numerical data to categorical
bank['income'] = pd.qcut(bank['income'], 4, labels=['very_low_income','low_income','medium_income','high_income'])
bank['age'] = pd.qcut(bank['age'], 3, labels=['young','middle_age','old'])

# replacing the values
bank['married'].replace({'NO': 'unmarried', 'YES': 'married'}, inplace=True)
bank['children'].replace({0:'no_children',1:'1_child',2:'2_children',3:'3_children'}, inplace=True)
bank['car'].replace({'NO': 'no_car', 'YES': 'car'}, inplace=True)
bank['save_act'].replace({'NO': 'no_savings_act', 'YES': 'savings_act'}, inplace=True)
bank['current_act'].replace({'NO': 'no_act', 'YES': 'act'}, inplace=True)
bank['mortgage'].replace({'NO': 'no_mortgage', 'YES': 'mortgage'}, inplace=True)
bank['pep'].replace({'NO': 'no_pep', 'YES': 'pep'}, inplace=True)
```

IST 707: Applied Machine Learning

Name: Chaithra Kopparam Cheluvaiah

SUID: 326926205

Email: ckoppara@syr.edu

RULE MINING

5. ASSOCIATION RULES & PEP

```
rules = list(apriori(bank.values,
        min_support=0.05,
        min_confidence=0.8,
        min_lift=2,
        max_length=None))
len(rules)
```

```
107
```

```
for rule in rules:
    stat = rule.ordered_statistics[0]
    if (list(stat.items_add)[0] not in ['yes_pep','no_pep']) or len(stat.items_add) > 1:
        continue

    print('Rule: ', list(stat.items_base), ' -> ', list(stat.items_add))
    print("Support: " + str(rule[1]))

    print("Confidence: " + str(stat.confidence))
    print("Lift: " + str(stat.lift))
    print("-------------------------------------")
```

IST 707: Applied Machine Learning

Name: Chaithra Kopparam Cheluvaiah

SUID: 326926205

Email: ckoppara@syr.edu

Experimenting with Support, Confidence, and Life

Support	Confidence	Lift	# of Rules Mined	Rules Mined
0.2	1	2	0	None
0.2	0.2	2	0	None
0.2	0.2	1.8	0	None
0.01	0.01	1.8	NA	computation was taking long time. Had to stop the process
0.5	0.5	1.8	0	None
0.02	0.6	2	2660	could not print all the rules
0.02	0.7	2	2340	could not print all the rules
0.08	0.8	3	1	['yes_accnt', 'yes_savings_act', 'yes_pep', 'old'] -> ['high_income']
0.05	0.8	3	7	['low_income', '1_child'] -> ['married', 'yes_pep']
				['no_mortgage', 'FEMALE', 'high_income'] -> ['yes_savings_act', 'old']
				['FEMALE', 'high_income', 'yes_pep'] -> ['yes_accnt', 'old']
				['FEMALE', 'high_income', 'yes_pep'] -> ['yes_savings_act', 'old']
				['married', 'high_income', 'yes_pep'] -> ['yes_savings_act', 'old']
				['yes_accnt', 'yes_savings_act', 'yes_pep', 'old'] -> ['high_income']
				['yes_savings_act', 'no_car', 'yes_accnt', 'no_mortgage', 'old'] -> ['high_income']

IST 707: Applied Machine Learning

Name: Chaithra Kopparam Cheluvaiah

SUID: 326926205

Email: ckoppara@syr.edu

Interesting rules that found while experimenting with support, confidence, and lift values:

1)

Support	Confidence	Lift	# of Rules Mined	Rules Mined
0.08	0.8	3	1	['yes_accnt', 'yes_savings_act', 'yes_pep', 'old'] -> ['high_income']

Support	Confidence	Lift	# of Rules Mined	Rules Mined
0.05	0.8	3	7	['low_income', '1_child'] -> ['married', 'yes_pep']
				['no_mortgage', 'FEMALE', 'high_income'] -> ['yes_savings_act', 'old']
				['FEMALE', 'high_income', 'yes_pep'] -> ['yes_accnt', 'old']
				['FEMALE', 'high_income', 'yes_pep'] -> ['yes_savings_act', 'old']
				['married', 'high_income', 'yes_pep'] -> ['yes_savings_act', 'old']
				['yes_accnt', 'yes_savings_act', 'yes_pep', 'old'] -> ['high_income']
				['yes_savings_act', 'no_car', 'yes_accnt', 'no_mortgage', 'old'] -> ['high_income']

From the above rules, we can infer that all the old age group people are having high incomes.

Name: Chaithra Kopparam Cheluvaiah

SUID: 326926205

Email: ckoppara@syr.edu

2) With 'PEP' on the RHS:

Support	Confidence	Lift	# of Rules Mined	Rules Mined
0.05	0.8	2	17	['low_income', '1_child'] -> ['yes_pep']
				['1_child', 'medium_income'] -> ['yes_pep']
				['1_child', 'middle_age'] -> ['yes_pep']
				['1_child', 'old'] -> ['yes_pep']
				['2_children', 'high_income'] -> ['yes_pep']
				['yes_accnt', '1_child', 'middle_age'] -> ['yes_pep']
				['1_child', 'yes_savings_act', 'middle_age'] -> ['yes_pep']
				['1_child', 'no_mortgage', 'old'] -> ['yes_pep']
				['1_child', 'yes_savings_act', 'old'] -> ['yes_pep']
				['no_mortgage', 'high_income', 'unmarried'] -> ['yes_pep']
				['no_mortgage', 'unmarried', 'no_children'] -> ['yes_pep']
				['yes_mortgage', 'no_savings_act', 'no_children'] -> ['yes_pep']
				['yes_accnt', '1_child', 'married', 'yes_savings_act'] -> ['yes_pep']
				['yes_accnt', 'no_mortgage', 'unmarried', 'no_children'] -> ['yes_pep']
				['yes_savings_act', 'no_mortgage', 'unmarried', 'no_children'] -> ['yes_pep']

Married people having atleast one children are signing up for PEP plan even if they have low income. On the other hand, unmarried people with only high income are signing up for PEP plan.

Company should target married people who have more liabilities and unmarried individuals with high income for sending PEP advertisement emails.