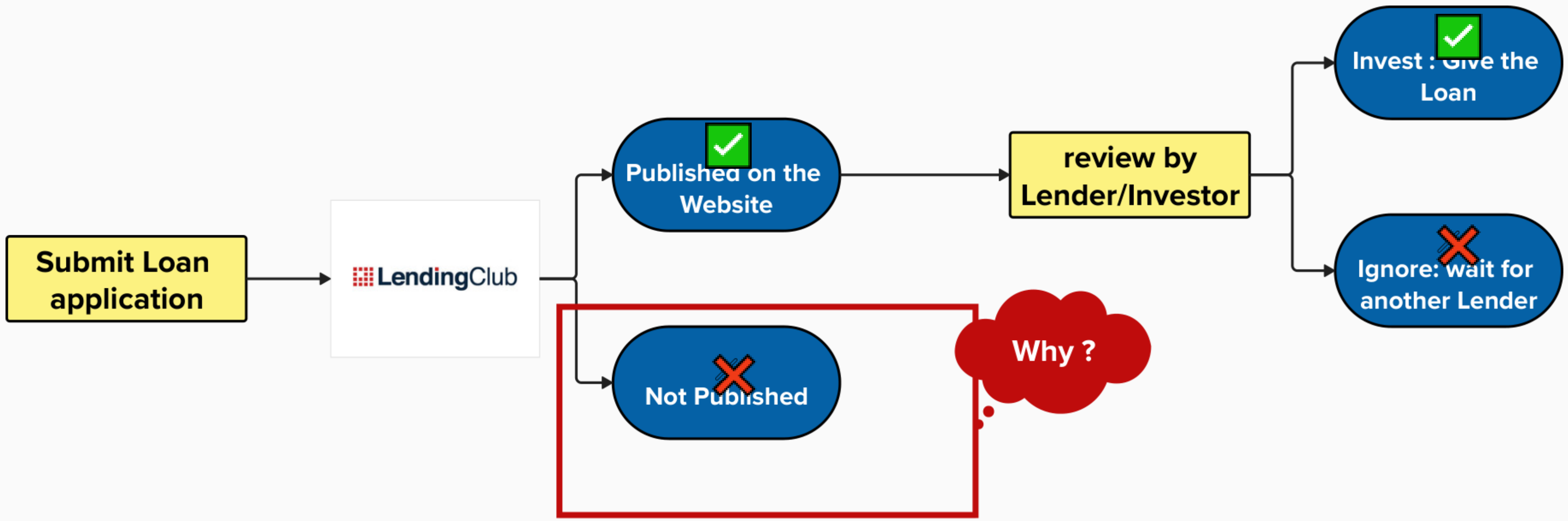
Understanding Rejected Loan Application on Lending Club Data

By Charles K

LOAN APPLICATION PROCESS



DATASET

NATHAN GEORGE · UPDATED 6 YEARS AGO

All Lending Club loan data

2007 through current Lending Club accepted and rejected loan data

kaggle.com/datasets/wordsforthewise/lending-club

GZIP

FEATURES	METRICS
amount_requested	US Dollars
application_date	YYYY-MM-DD
loan_title	Txt
risk_score	Int
debt_to_income_ratio	Pourcentage
zip_code	Code
state	Code
employment_length	Txt
policy_code	Int

Project Phase

PROJECT PHASE	COMMENTS	TECHNIQUES
Step 1 : Exploratory Data Analysis	Load the Dataset	*Read .gzip with Pandas on Google Colab*
	Understand the Metrics	*Matplotlib*
	Convert the features	*Python Built-in functions*
	Delete useless rows	*Python Built-in functions*
Step 2 : Preprocess the Data	Delete useless Features	*Python Built-in functions*
	Correlation matrix	*.corr Pandas function*
	Impute missing values on employment_length and policy_code	*Python Built-in functions*
	Normalise 'loan_title'	*Regular Expression and Manual regrouping*
	Scale the features	*.StandardScaler Sklearn*
	Encode 'state'	*.OneHotEncoder SKlearn*
	Transform loan_title to numeric	*TF-IDF Sklearn*
Step 3 : Building the Model	Topic Modeling on 'loan_title'	*Non-Negative Matrix Factorization (NMF)*
	Clustering	*K-Means++*
Step 4 : Conclusion and Data Viz	Visualize the Topics	*Principal Component Analysis (PCA)*
	Result of the Toic Modeling	*Centroids, and Summary*

KEY CONCEPTS

TF-IDF

OneHotEncoder

Non-Negative Matrix Factorization (NMF)

K-Means++

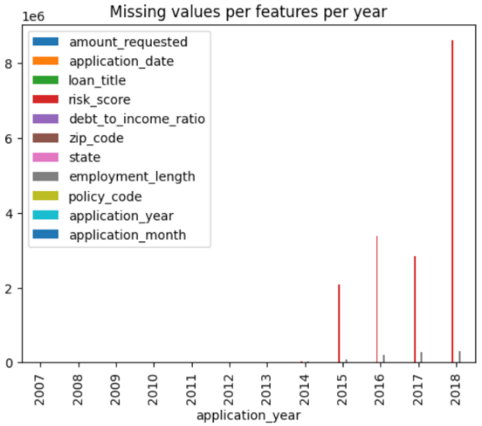
Principal Component Analysis (PCA)

EDA & PRE-PROCESS 1/6

```
Sum of Null values:
 amount_requested           0
application_date            0
loan_title               1305
risk_score           18497630
debt_to_income_ratio        0
zip_code                  293
state                      22
employment_length      951355
policy_code               918
```

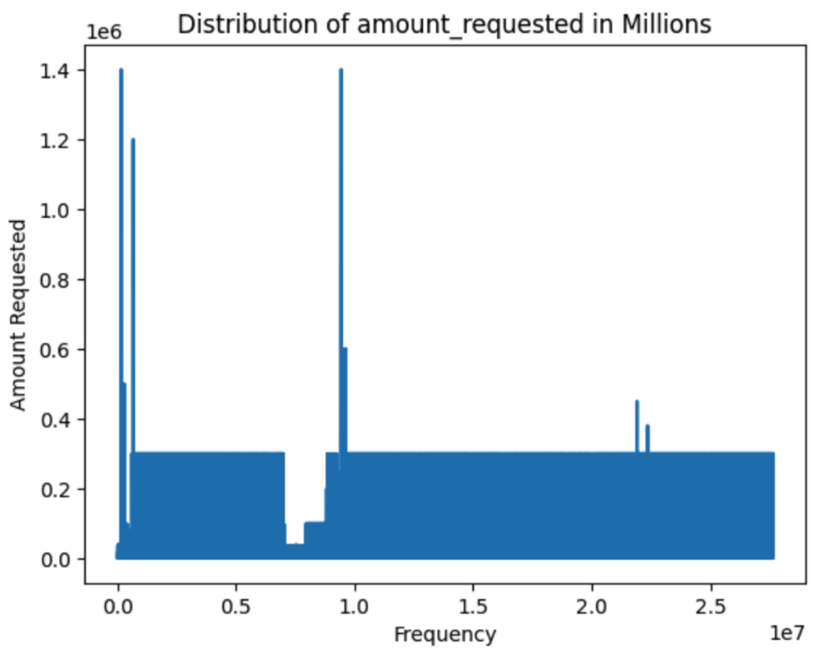
```
 #   Column                Dtype
---  ------                -----
 0   amount_requested      float64
 1   application_date      object
 2   loan_title            object
 3   risk_score            float64
 4   debt_to_income_ratio  object
 5   zip_code              object
 6   state                 object
 7   employment_length     object
 8   policy_code           float64
```

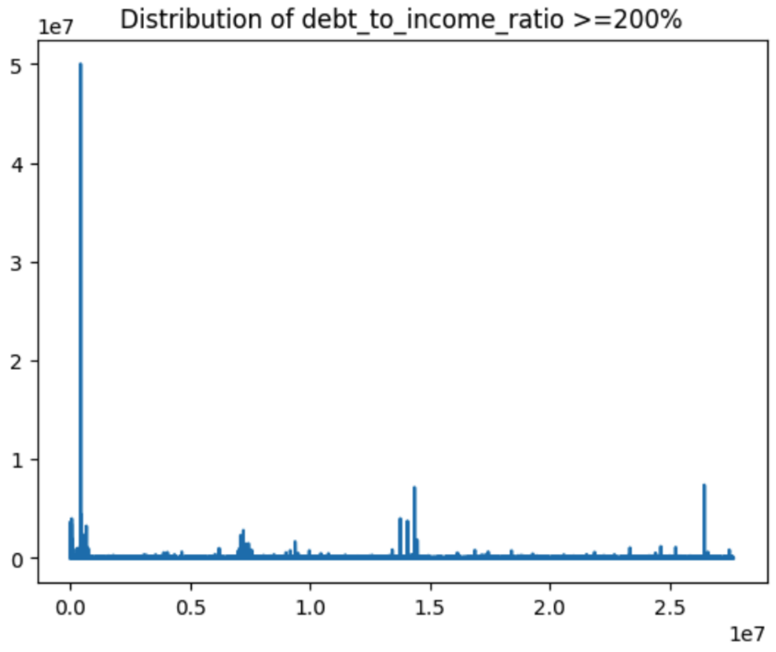
```
Description of numeric values:
        amount_requested    risk_score   policy_code
count       2.764874e+07  9.151111e+06  2.764782e+07
mean        1.313324e+04  6.281721e+02  6.375113e-03
std         1.500964e+04  8.993679e+01  1.127368e-01
min         0.000000e+00  0.000000e+00  0.000000e+00
25%         4.800000e+03  5.910000e+02  0.000000e+00
50%         1.000000e+04  6.370000e+02  0.000000e+00
75%         2.000000e+04  6.750000e+02  0.000000e+00
max         1.400000e+06  9.900000e+02  2.000000e+00
```



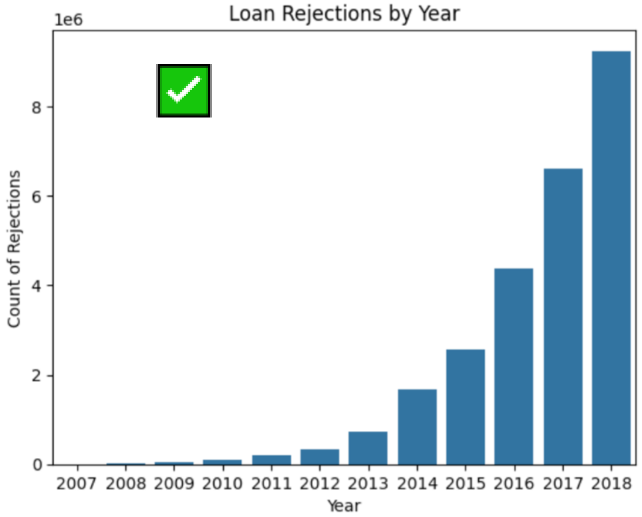
- 27 Millions rejected loans
- Max of $1.4 Millions amount_requested
- Empty values on amount_requested
- Empty loan_title
- 18 Millions (67%) missing risk_score
- 5% missing values on employment_length
- Debt_to_income & state in bad format

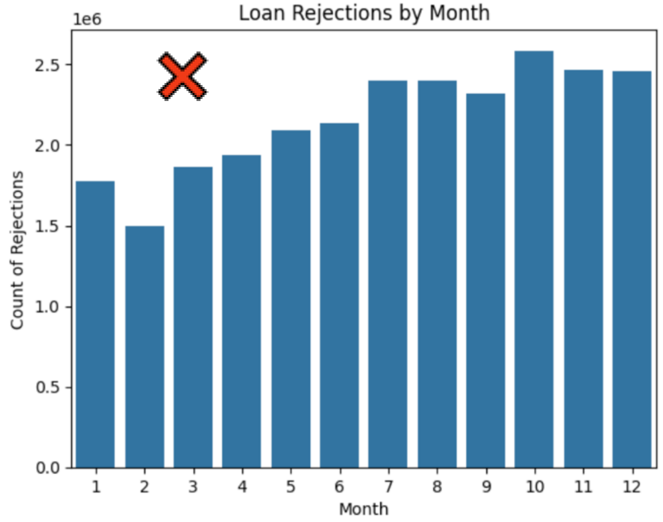
EDA & PRE-PROCESS 2/6

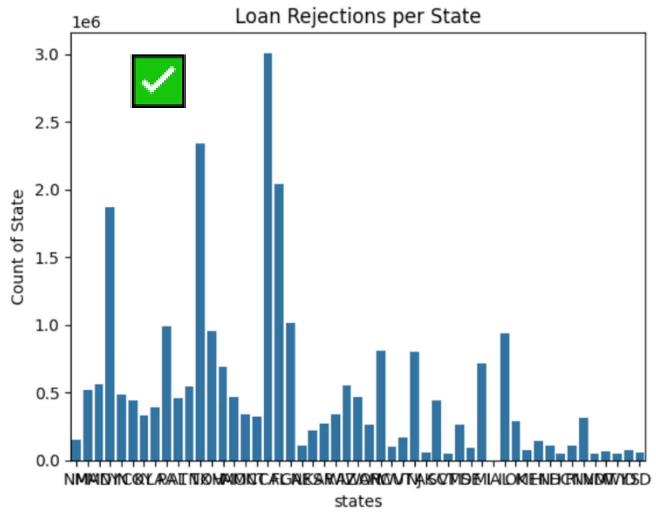




EDA & PRE-PROCESS 3/6

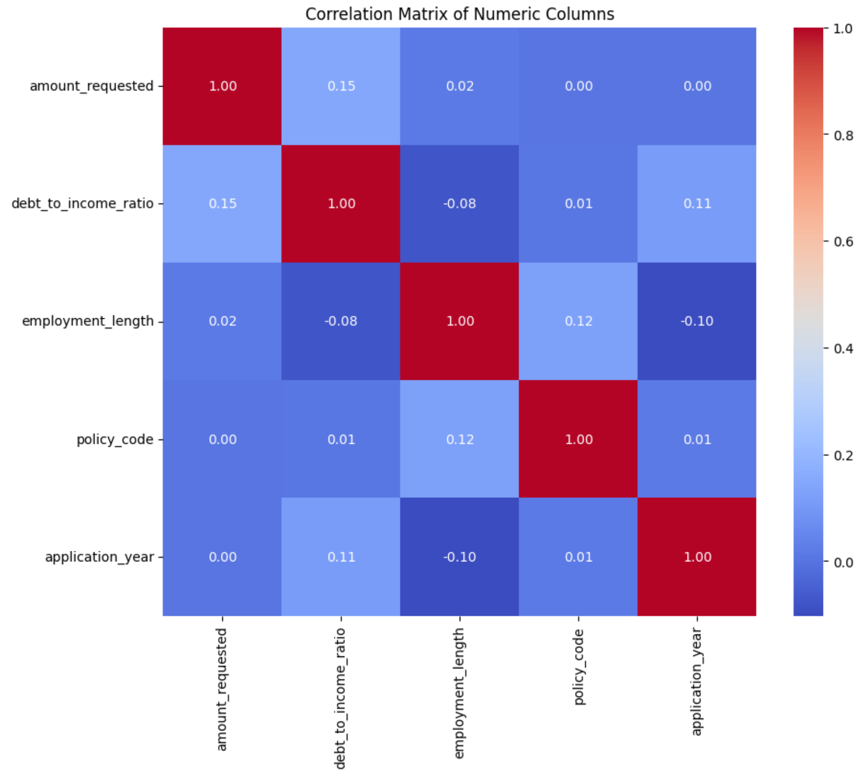






EDA & PRE-PROCESS 4/6

FINAL FEATURES
amount_requested
debt_to_income_ratio
employment_length
policy_code
application_year
loan_title_grouped (loan_title)



EDA & PRE-PROCESS 5/6

TECHNIQUES	FEATURES
Filling missing value : Media imputation	employment_length
Standar Scaler	amount_requested debt_to_income_ratio employment_length
One-hot encoder	State
IF-IDF	loan_title_grouped (loan_title)

EDA & PRE-PROCESS 6/6

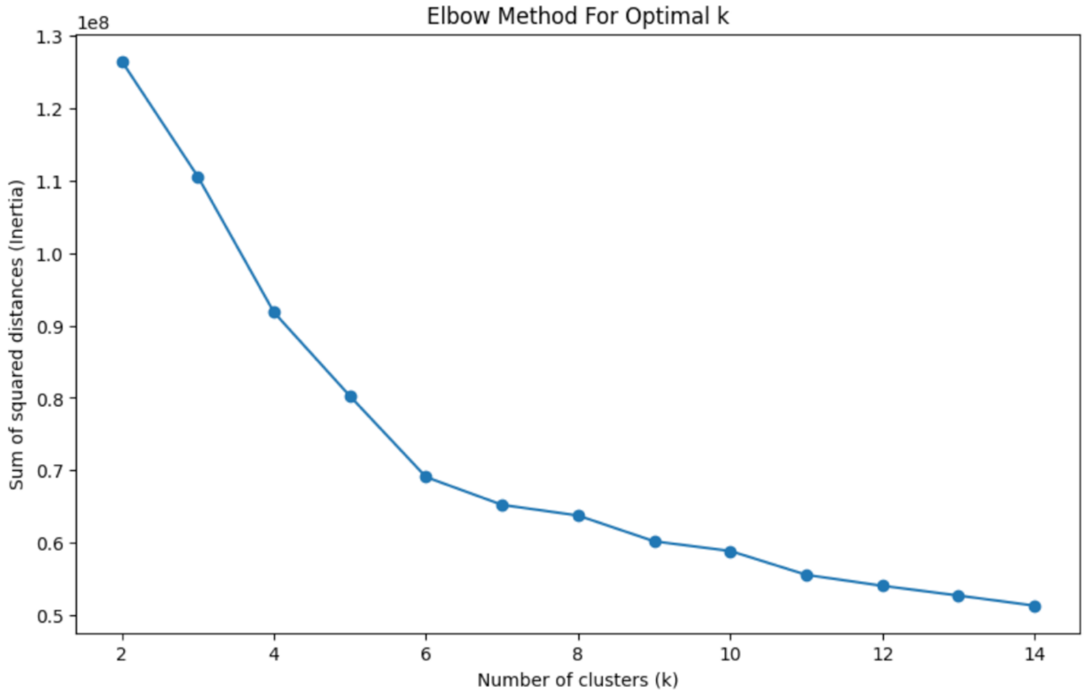
```
loan_title
debt consolidation            6296497
debtconsolidation             5461852
other                         4443147
credit card refinancing       2248845
creditcard                    1259856
car financing                  751178
homeimprovement                655212
home buying                    485627
major purchase                 474369
car                            468824
home improvement               457402
majorpurchase                  425006
medical expenses               381935
moving                         338413
medical                        324225
moving and relocation          311483
vacation                       282705
smallbusiness                  280819
business                       208786
house                          165065
renewableenergy                 27752
green loan                      23436
wedding                         17701
consolidation                    3183
personal loan                    3107
personal                         2459
debt consolidation loan          2106
educational                      2012
student loan                     1258
credit card consolidation        1187
Name: count, dtype: int64
```

```
loan_title_grouped
Debt Consolidation       11771139
Other                     4444090
Credit Card               3516496
Car Financing             1221367
Home Improvement          1113817
Major Purchase             899375
Medical                    706348
Home Buying                650981
Moving and Relocation      650160
Small Business             492037
Vacation                   282705
Green Loan                  51199
Wedding                     18188
Personal Loan                7818
Educational Loan             3954
Student Loan                 1258
Freedom                       548
Pool Loan                     296
New Start                     279
major purchase loan           171
Name: count, dtype: int64
```

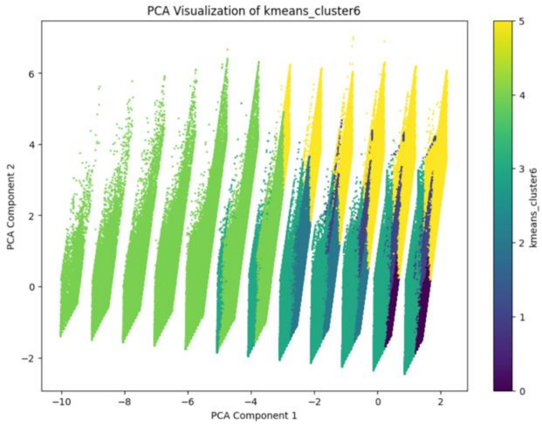
TOPIC MODELING: NMF

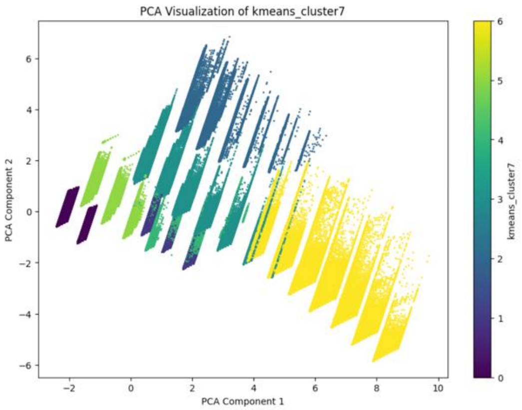
```
Topic 0:
consolidation debt pay loan consolidate free help consolidating paying payoff personal cc
Topic 1:
card credit pay cards payoff consolidate refinance loan high paying payment help
Topic 2:
improvement home loan payment new project repair improvment repairs purchase consolidate help
Topic 3:
financing car need new repair used repairs pay loan purchase payment buy
Topic 4:
major purchase loan car home motorcycle equipment vehicle inventory auto business new
Topic 5:
medical bills expenses pay expense help school loan need consolidate payoff cards
Topic 6:
relocation moving expenses loan forward expense job new help school need home
Topic 7:
small business start loan startup new expansion capital investment starting expanding expand
Topic 8:
buying home payment new repair improvment repairs purchase loan help car need
Topic 9:
vacation loan family dream summer money time pay needed need hawaii home
Topic 10:
loan green personal educational student pay pool need payoff school refinance cc
Topic 11:
wedding expenses expense dream help daughters loan fund day pay ring sons
```

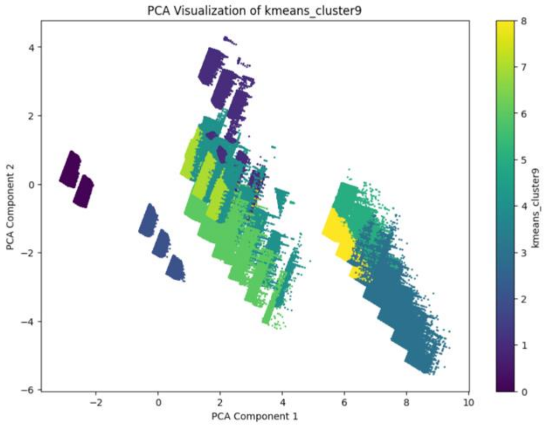
CLUSTERING : K-MEANS

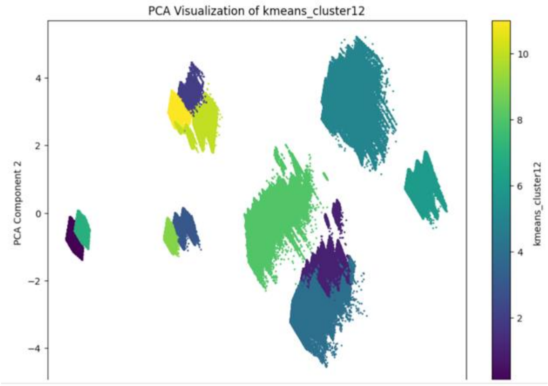


CLUSTERING : K-MEANS





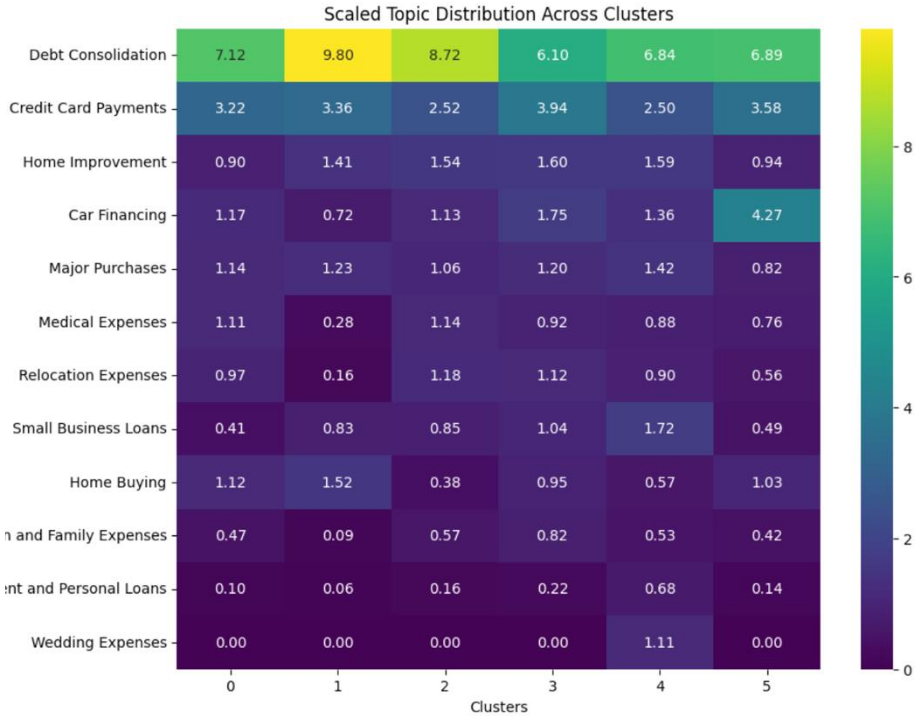


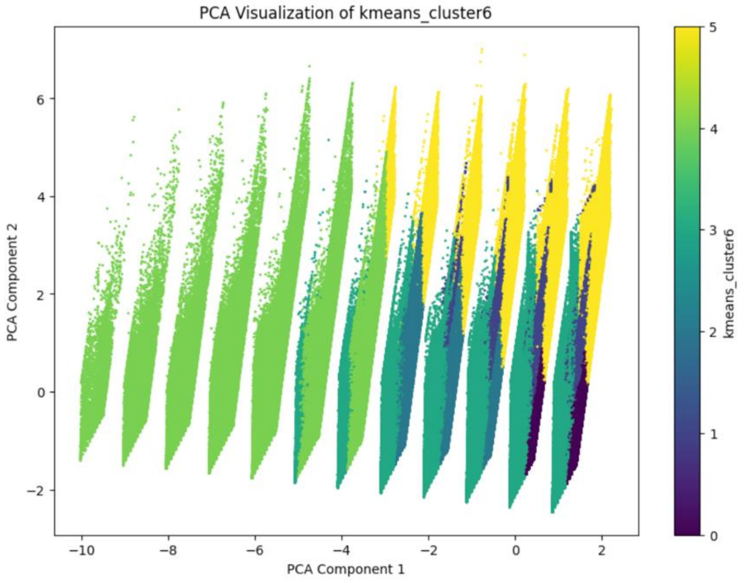


SILHOUETTE SCORE	
k=6	0.25
k=7	0.26
k=9	0.10
k=12	0.13

VISUALISATION

Topic_0	Debt Consolidation
Topic_1	Credit Card Payments
Topic_2	Home Improvement
Topic_3	Car Financing
Topic_4	Major Purchases
Topic_5	Medical Expenses
Topic_6	Relocation Expenses
Topic_7	Small Business Loans
Topic_8	Home Buying
Topic_9	Vacation and Family Expenses
Topic_10	Student and Personal Loans
Topic_11	Wedding Expenses





CONCLUSION : TAKEAWAY