US University Students' Borrowings

Eldo Martadjaja

Suggested solution

A Machine Learning approach to predict the number of loan borrowers who completed college based on their economy and education

[redacted]

Agenda

- The data set
- Use case introduction
- The solution

[redacted]

Data Set

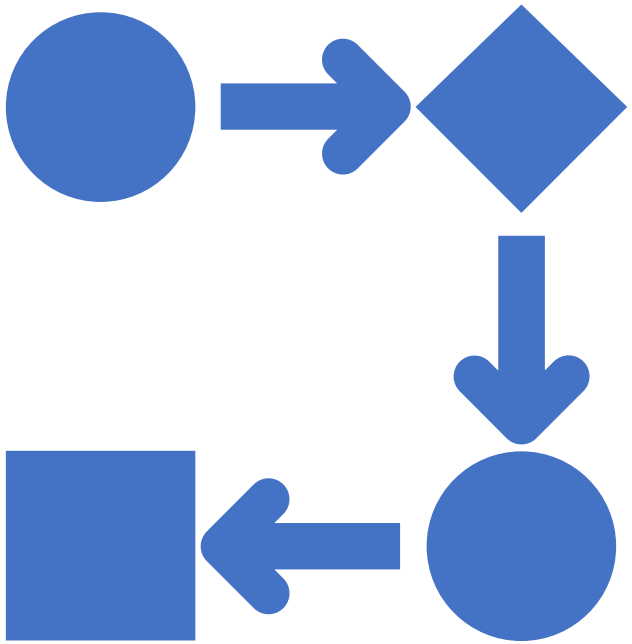
- ❑US Department of Education (https://collegescorecard.ed.gov/data/)
- ❑College Scorecard student data
- ❑216638 entries

Data Dictionary

1. UNITID: Unit ID for institution
2. OPEID6: 6-digit OPE ID for institution
3. INSTNM: Institution name
4. CONTROL: Control of institution
5. MAIN: Flag for main campus
6. CIPCODE:Classification of Instructional Programs (CIP) code for the field of study
7. CIPDESC: Text description of the field of study CIP Code
8. CREDLEV: Level of credential
9. CREDDESC:Text description of the level of credential
10. COUNT: Number of borrowers of federal loans completing college
11. DEBTMEDIAN: Median federal loan debt of borrowers completing college
12. DEBTPAYMENT10YR: Median federal loan debt of borrowers completing college in m
13. DEBTMEAN: Mean federal loan debt of borrowers completing college
14. TITLEIVCOUNT: Number of federally-aided students completing college
15. EARNINGSCOUNT: Number of federally-aided students completing
16. MD_EARN_WNE: Median earnings of federally-aided completers i
17. IPEDSCOUNT1: Number of awards to all students in year 1 of the
18. IPEDSCOUNT2: Number of awards to all students in year 2 of the

The use case

Exploration of student’s tendency, who borrowed money through loans, to complete college based on the field of study, level of education and other factors



Workflow

[redacted]

Exploring US College Students' trends in their economy based on their field of study

For this project, I have obtained the dataset from the US Department of Education (https://collegescorecard.ed.gov/data/), which contains college student data including subject, earnings, debts, payments etc. For that reason, I decided to explore students' trend by subject and look through their incomes and expenses.

Data Dictionary

1. UNITID: Unit ID for institution
2. OPEID6: 6-digit OPE ID for institution
3. INSTNM: Institution name
4. CONTROL: Control of institution
5. MAIN: Flag for main campus
6. CIPCODE:Classification of Instructional Programs (CIP) code for the field of study
7. CIPDESC: Text description of the field of study CIP Code
8. CREDLEV: Level of credential
9. CREDDESC:Text description of the level of credential
10. COUNT: Number of borrowers of federal loans completing college
11. DEBTMEDIAN: Median federal loan debt of borrowers completing college
12. DEBTPAYMENT10YR: Median federal loan debt of borrowers completing college in monthly payments (10-year amortization pla[redacted]
13. DEBTMEAN: Mean federal loan debt of borrowers completing college
14. TITLEIVCOUNT: Number of federally-aided students completing college
15. EARNINGSCOUNT: Number of federally-aided students completing college in the earnings cohort
16. MD_EARN_WNE: Median earnings of federally-aided completers in the earnings cohort
17. IPEDSCOUNT1: Number of awards to all students in year 1 of the pooled debt cohort
18. IPEDSCOUNT2: Number of awards to all students in year 2 of the pooled debt cohort

Data Science Peers Presentation

Contents

Data Dictionary

1. UNITID: Unit ID for institution
2. OPEID6: 6-digit OPE ID for institution
3. INSTNM: Institution name
4. CONTROL: Control of institution
5. MAIN: Flag for main campus
6. CIPCODE:Classification of Instructional Programs (CIP) code for the field of study
7. CIPDESC: Text description of the field of study CIP Code
8. CREDLEV: Level of credential
9. CREDDESC:Text description of the level of credential
10. COUNT: Number of borrowers of federal loans completing college
11. DEBTMEDIAN: Median federal loan debt of borrowers completing college
12. DEBTPAYMENT10YR: Median federal loan debt of borrowers completing college in monthly p
13. DEBTMEAN: Mean federal loan debt of borrowers completing college
14. TITLEIVCOUNT: Number of federally-aided students completing college
15. EARNINGSCOUNT: Number of federally-aided students completing college in the earnings co
16. MD_EARN_WNE: Median earnings of federally-aided completers in the earnings cohort
17. IPEDSCOUNT1: Number of awards to all students in year 1 of the pooled debt cohort
18. IPEDSCOUNT2: Number of awards to all students in year 2 of the pooled debt cohort

Architecture

Supervised Machine Learning Algorithm – Labele[redacted]

Data quality assessment, data pre-processing and feature engineering

Model performance indicators

Coefficient of determination (R^2)

Model algorithm

TARGET VARIABLE: COUNT – REGRESSION PROBLEM

LINEAR REGRESSION PROVIDED BY SCIKIT LEARN

FEATURE SELECTION TO IDENTIFY KEY

Results – Model Performance

```
reg.fit(Xtrain, ytrain)

print('The training score is {}'.format(reg.score(Xtrain, ytrain)))
```

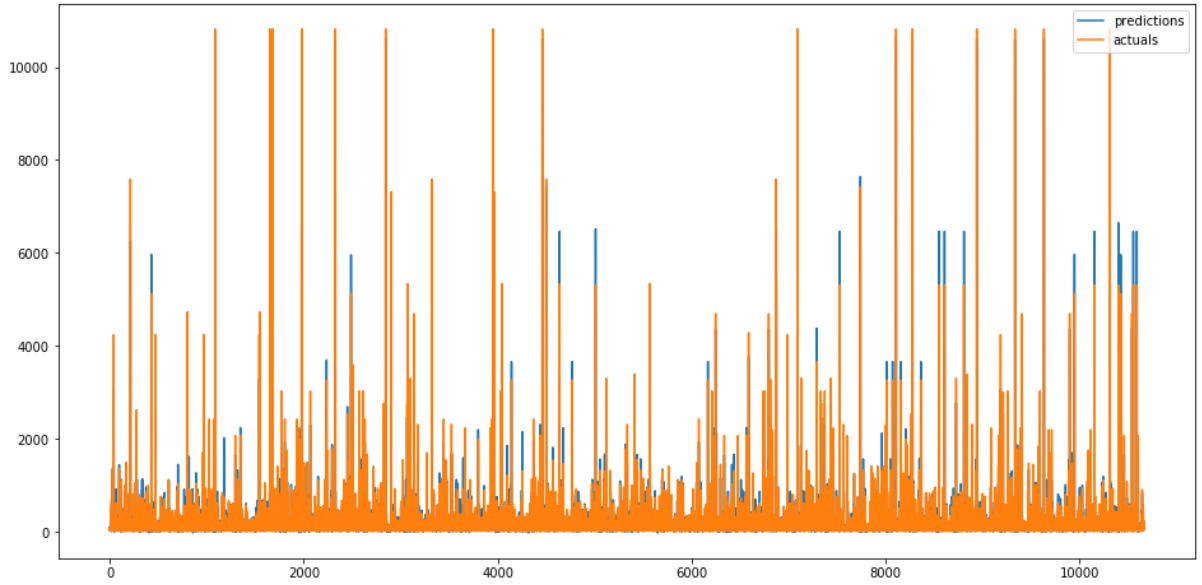
```
The training score is 0.9621833506154973
```

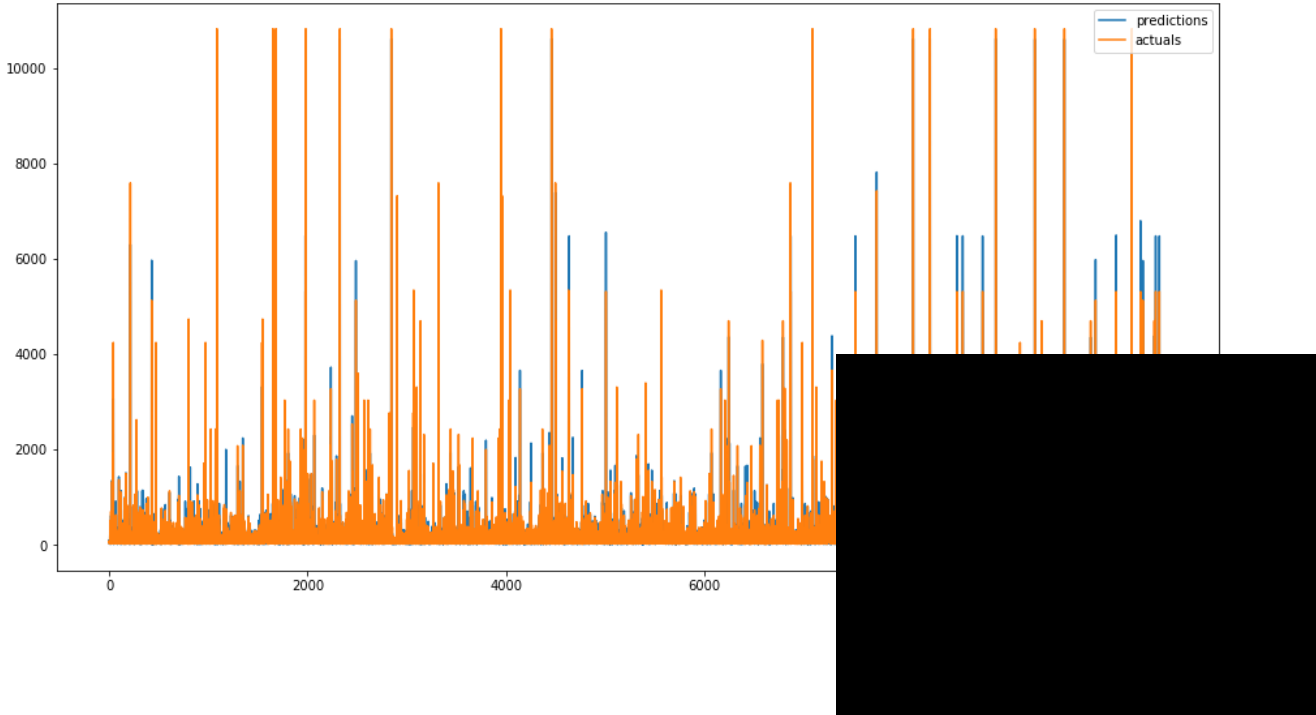
```
print('The evaluation R2 is {}'.format(r2_score(ytest, preds)))
```

```
The evaluation R2 is 0.963330003820413
```

```
print('The R2 score on the test data after training with {} features is {}%'
      .format(len(indices),r2_score(ytest, predictions)))
```

```
The R2 score on the test data after training with 4 features is 0.9625727857132932%
```

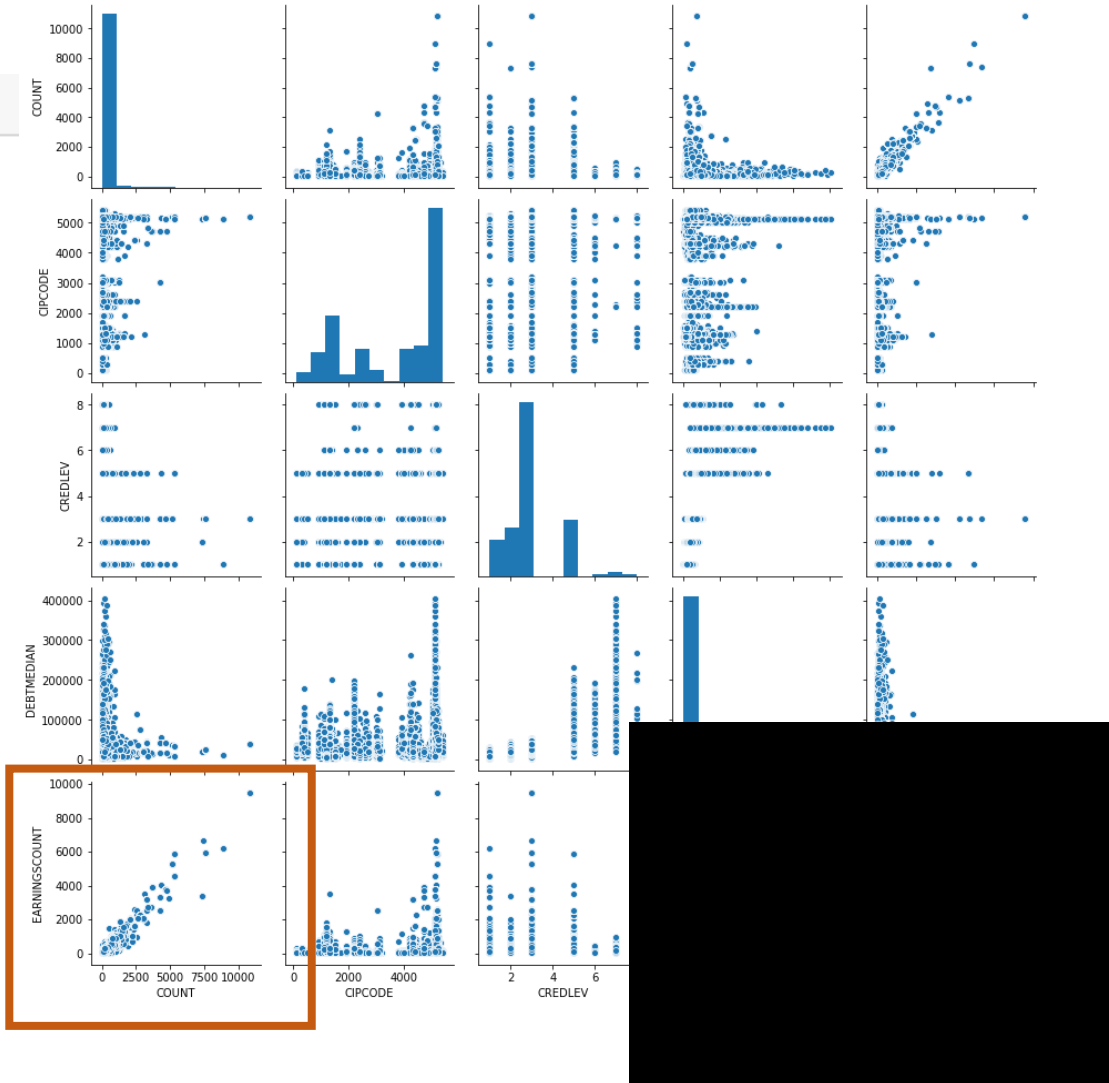




Results - Correlation

```
df.iloc[:, indices2].corr().sort_values(by='COUNT', ascending = False)
```

	COUNT	CIPCODE	CREDLEV	DEBTMEDIAN	EARNINGSCOUNT
COUNT	1.000000	0.086662	-0.043430	0.038674	0.978708
EARNINGSCOUNT	0.978708	0.084134	-0.026392	0.046449	1.000000
CIPCODE	0.086662	1.000000	-0.040325	0.070859	0.084134
DEBTMEDIAN	0.038674	0.070859	0.642103	1.000000	0.046449
CREDLEV	-0.043430	-0.040325	1.000000	0.642103	-0.026392



Thank you for your time