

Recommender Systems for Higher Education

Author: Hussain Abbas, MSc

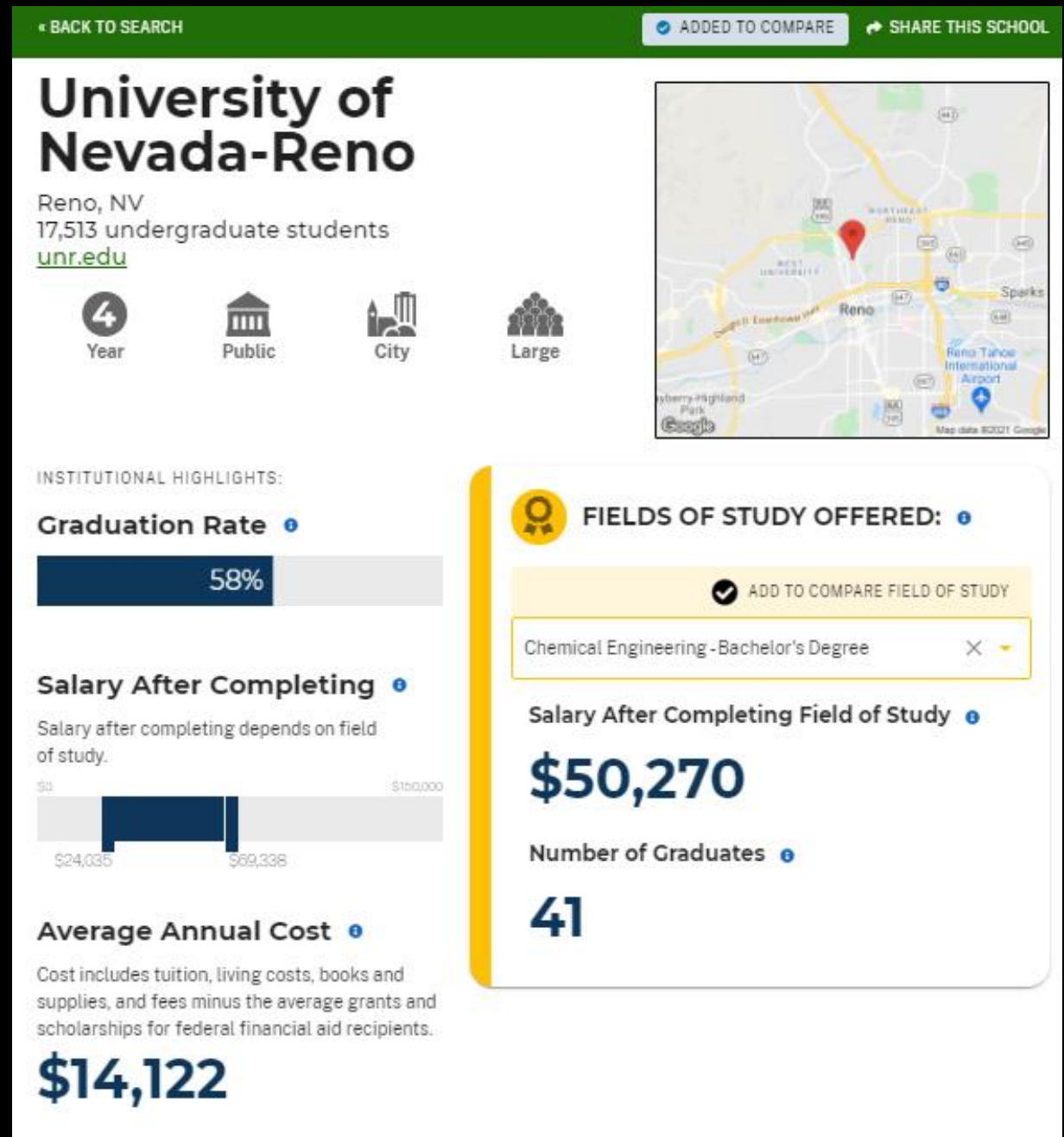
© 2021 Stats AI LLC

All Rights Reserved

The College Scorecard

- “Graduation Rate” and “Salary after Completing” are at the Institution Level, not at the Program level
- Some programs don’t have salary data due to small sample sizes

Stats AI



FIND THE RIGHT FIT.

Search and compare colleges: their fields of study, costs, admissions, results, and more.

SEARCH SCHOOLS

SEARCH FIELDS OF STUDY

SHOW ME OPTIONS

Select one or more options to create a list of schools that fit your needs.



Schools Near Me



I Want a Certificate



Most People Get In ⓘ



I Want an Associate's Degree



Most People Graduate ⓘ



I Want a Bachelor's Degree

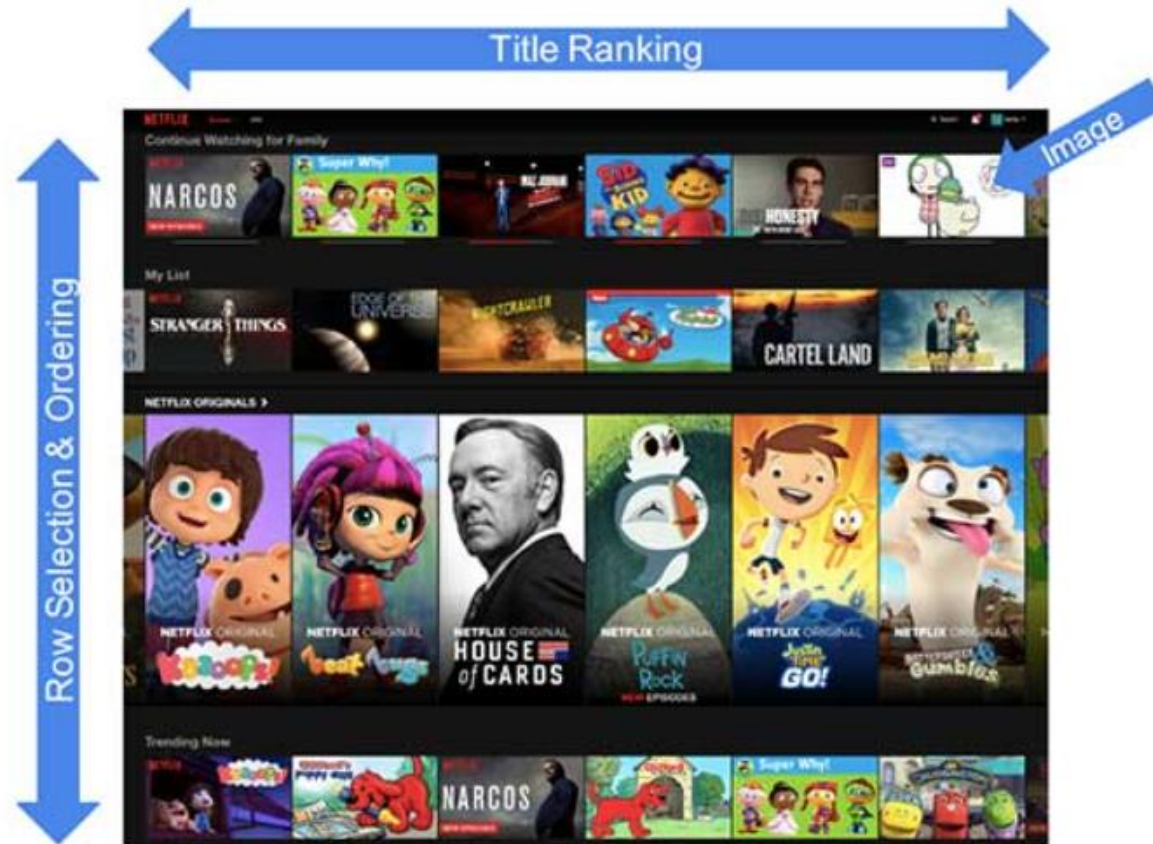
FIND SCHOOLS

- And to get there we have to “search” for it ...

- Does Netflix make you search for movies?
- Does Youtube make you search for videos?
- While both provide search, the value add is in ...

Recommender Systems!

Everything is a Recommendation



NETFLIX

Recommendations are driven by machine learning algorithms

Over 80% of what members watch comes from our recommendations

Stats AI

Motivation

- Users have a finite amount of time to spend on your website
- Instead of spending time searching for content, users should spend that time consuming the content
- Thus, RS enable users to skip the search process and jump straight to the content

FIND THE RIGHT FIT.

Search and compare colleges: their fields of study, costs, admissions, results, and more.

SEARCH SCHOOLS

SEARCH FIELDS OF STUDY

SHOW ME OPTIONS

Select one or more options to create a list of schools that fit your needs.



Schools Near Me



I Want a Certificate



Most People Get In ⓘ



I Want an Associate's Degree



Most People Graduate ⓘ



I Want a Bachelor's Degree

FIND SCHOOLS

- So why is this still a thing? 🤔🤔🤔

Benefits of framing the problem as a RS Task

- We don't have to worry about finding the best metric that values a given program
- Instead, the user tells us what's important to them
- Or we just grab it from their Facebook profile ...
- The best RS is the one that users engage with the most

The College Scorecard website only shows a snippet of the available data!

Data Documentation

College Scorecard provides data at the institution-level and data by field of study. The technical data documents provide in-depth information about these data. The data dictionary includes separate worksheets with institution-level and field of study dictionaries as well as cohort maps describing the timing aspects of each data element.

Data Dictionary

**Technical Documentation
for Institution-Level Data
Files**

**Technical Documentation
for Data Files by Field of
Study**

Source: <https://collegescorecard.ed.gov/data/documentation/>

CollegeScorecardDataDictionary - Excel					Sign in
File Home Insert Page Layout Formulas Data Review View Help					Share Comment
A32				Median Parent PLUS loan debt disbursed to non-males at all institutions	
1	NAME OF DATA ELEMENT	dev-categ	developer-friendly name	API data type	VARIABLE NAME
20					
21	Text description of the level of credential	programs	cip_4_digit.credential.title	string	CREDDESC
22	Number of awards to all students in year 1 of the pooled debt cohort	programs	cip_4_digit.counts.ipeds_awards1	integer	IPEDSCOUNT1
23	Number of awards to all students in year 2 of the pooled debt cohort	programs	cip_4_digit.counts.ipeds_awards2	integer	IPEDSCOUNT2
24	Student recipient count for average/median Parent PLUS loan debt disbursed at all institutions	programs	cip_4_digit.debt.parent_plus.all.all_inst.count	integer	DEBT_ALL_PP_ANY_N
25	Average Parent PLUS loan debt disbursed at all institutions	programs	cip_4_digit.debt.parent_plus.all.all_inst.average	integer	DEBT_ALL_PP_ANY_MEAN
26	Median Parent PLUS loan debt disbursed at all institutions	programs	cip_4_digit.debt.parent_plus.all.all_inst.median	integer	DEBT_ALL_PP_ANY_MDN
27	Student recipient count for average/median Parent PLUS loan debt disbursed to males at all institutions	programs	cip_4_digit.debt.parent_plus.male.all_inst.count	integer	DEBT_MALE_PP_ANY_N
28	Average Parent PLUS loan debt disbursed to males at all institutions	programs	cip_4_digit.debt.parent_plus.male.all_inst.average	integer	DEBT_MALE_PP_ANY_MEAN
29	Median Parent PLUS loan debt disbursed to males at all institutions	programs	cip_4_digit.debt.parent_plus.male.all_inst.median	integer	DEBT_MALE_PP_ANY_MDN
30	Student recipient count for average/median Parent PLUS loan debt disbursed to non-males at all institutions	programs	cip_4_digit.debt.parent_plus.not_male.all_inst.count	integer	DEBT_NOTMALE_PP_ANY_N
31	Average Parent PLUS loan debt disbursed to non-males at all institutions	programs	cip_4_digit.debt.parent_plus.not_male.all_inst.average	integer	DEBT_NOTMALE_PP_ANY_MEAN
32	Median Parent PLUS loan debt disbursed to non-males at all institutions	programs	cip_4_digit.debt.parent_plus.not_male.all_inst.median	integer	DEBT_NOTMALE_PP_ANY_MDN
33	Student recipient count for average/median Parent PLUS loan debt disbursed to Pell recipients at all institutions	programs	cip_4_digit.debt.parent_plus.pell.all_inst.count	integer	DEBT_PELL_PP_ANY_N
34	Average Parent PLUS loan debt disbursed to Pell recipients at all institutions	programs	cip_4_digit.debt.parent_plus.pell.all_inst.average	integer	DEBT_PELL_PP_ANY_MEAN
35	Median Parent PLUS loan debt disbursed to Pell recipients at all institutions	programs	cip_4_digit.debt.parent_plus.pell.all_inst.median	integer	DEBT_PELL_PP_ANY_MDN
36	Student recipient count for average/median Parent PLUS loan debt disbursed to non-Pell-recipients at all institutions	programs	cip_4_digit.debt.parent_plus.not_pell.all_inst.count	integer	DEBT_NOPELL_PP_ANY_N
37	Average Parent PLUS loan debt disbursed to non-Pell-recipients at all institutions	programs	cip_4_digit.debt.parent_plus.not_pell.all_inst.average	integer	DEBT_NOPELL_PP_ANY_MEAN
38	Median Parent PLUS loan debt disbursed to non-Pell-recipients at all institutions	programs	cip_4_digit.debt.parent_plus.not_pell.all_inst.median	integer	DEBT_NOPELL_PP_ANY_MDN
39	Student recipient count for average/median Parent PLUS loan debt disbursed at this institution	programs	cip_4_digit.debt.parent_plus.all.eval_inst.count	integer	DEBT_ALL_PP_EVAL_N
40	Average Parent PLUS loan debt disbursed at this institution	programs	cip_4_digit.debt.parent_plus.all.eval_inst.average	integer	DEBT_ALL_PP_EVAL_MEAN
41	Median Parent PLUS loan debt disbursed at this institution	programs	cip_4_digit.debt.parent_plus.all.eval_inst.median	integer	DEBT_ALL_PP_EVAL_MDN
42	Student recipient count for average/median Parent PLUS loan debt disbursed to males at this institution	programs	cip_4_digit.debt.parent_plus.male.eval_inst.count	integer	DEBT_MALE_PP_EVAL_N

AutoSave Off										Most-Recent-Cohorts-Field-of-Study										Search										Sign in									
File Home Insert Page Layout Formulas Data Review View Help										Share										Comments																			
K2275										26																													
	A	B	C		D	E	F	G	H	I	J	K	L	M		N	O	P	Q																				
1	UNIT	OPEID6	INSTNM	CONTR	M	CIPCOD	CIPDESC	CREDLE	CREDDI	IPEDSC	IPEDSC	DEBT_A	DEBT_ALL_STGP_ANY_ME	DEBT_A	DEBT_A	DEBT_A	DEBT_A	DEBT_A	DEBT_A																				
355	100724	1005	Alabama State University	Public	1	1107	Computer Science.	3	Bachelors	4	11	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
660	100830	8310	Auburn University at Montgomery	Public	1	1107	Computer Science.	3	Bachelors	2	3	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
1108	101189	1003	Faulkner University	Private, no	1	1107	Computer Science.	3	Bachelors	0	3	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
1301	101365	9621	Herzing University-Birmingham	Private, no	0	1107	Computer Science.	3	Bachelors	NULL	NULL	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
1973	101912	1033	Oakwood University	Private, no	1	1107	Computer Science.	3	Bachelors	0	2	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
2098	102049	1036	Samford University	Private, no	1	1107	Computer Science.	3	Bachelors	7	0	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
2275	102094	1057	University of South Alabama	Public	1	1107	Computer Science.	3	Bachelors	24	26	30	30954	27500	29	29779	26000																						
2591	102377	1050	Tuskegee University	Private, no	1	1107	Computer Science.	3	Bachelors	1	6	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
2902	102614	1063	University of Alaska Fairbanks	Public	1	1107	Computer Science.	3	Bachelors	3	1	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
3303	103778	1081	Thunderbird School of Global Management	Public	0	1107	Computer Science.	3	Bachelors	NULL	NULL	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
3591	103945	31203	CollegeAmerica-Flagstaff	Private, no	1	1107	Computer Science.	3	Bachelors	1	NULL	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
3719	104151	1081	Arizona State University-Tempe	Public	1	1107	Computer Science.	3	Bachelors	292	324	255	21562	19281	250	20130	18750																						
4225	104179	1083	University of Arizona	Public	1	1107	Computer Science.	3	Bachelors	169	211	136	25863	25000	133	24307	24747																						
4898	104586	1479	Embry-Riddle Aeronautical University-Prescott	Private, no	0	1107	Computer Science.	3	Bachelors	NULL	NULL	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
5030	104717	1074	Grand Canyon University	Private, fo	1	1107	Computer Science.	3	Bachelors	0	13	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
5517	105330	1082	Northern Arizona University	Public	1	1107	Computer Science.	3	Bachelors	42	48	57	22782	23500	56	21327	24604																						
6652	106342	1088	Lyon College	Private, no	1	1107	Computer Science.	3	Bachelors	4	9	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
6708	106397	1108	University of Arkansas	Public	1	1107	Computer Science.	3	Bachelors	45	48	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
7361	106704	1092	University of Central Arkansas	Public	1	1107	Computer Science.	3	Bachelors	0	0	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
7696	107044	1097	Harding University	Private, no	1	1107	Computer Science.	3	Bachelors	26	33	29	PrivacySuppressed	24725	29	PrivacySup	24625																						
7855	107080	1099	Hendrix College	Private, no	1	1107	Computer Science.	3	Bachelors	13	7	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
8046	107512	1102	Ouachita Baptist University	Private, no	1	1107	Computer Science.	3	Bachelors	2	6	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
8224	107600	1103	Philander Smith College	Private, no	1	1107	Computer Science.	3	Bachelors	8	5	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
9424	110097	1122	Biola University	Private, no	1	1107	Computer Science.	3	Bachelors	17	21	30	PrivacySuppressed	24250	30	PrivacySup	24000																						
9775	110361	1125	California Baptist University	Private, no	1	1107	Computer Science.	3	Bachelors	1	1	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
9894	110404	1131	California Institute of Technology	Private, no	1	1107	Computer Science.	3	Bachelors	67	57	PrivacySup	PrivacySuppressed	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup	PrivacySup																				
9977	110413	1133	California Lutheran University	Private, no	1	1107	Computer Science.	3	Bachelors	34	17	20	19621	20811	19	19404	18500																						
10063	110422	1143	California Polytechnic State University-San Luis Obispo	Public	1	1107	Computer Science.	3	Bachelors	150	169	68	PrivacySuppressed	18500	68	PrivacySup	18500																						
10197	110486	7993	California State University-Bakersfield	Public	1	1107	Computer Science.	3	Bachelors	26	49	13	24840	PrivacySup	11	25283	PrivacySup																						
10261	110495	1157	California State University-Stanislaus	Public	1	1107	Computer Science.	3	Bachelors	41	62	51	PrivacySuppressed	10854	51	PrivacySup	10256																						
10315	110510	1142	California State University-San Bernardino	Public	1	1107	Computer Science.	3	Bachelors	51	69	57	18535	19092	55	17832	15903																						
10438	110529	1144	California State Polytechnic University-Pomona	Public	1	1107	Computer Science.	3	Bachelors	198	207	204	19284	20166	202	18795	20000																						
10526	110538	1146	California State University-Chico	Public	1	1107	Computer Science.	3	Bachelors	43	50	42	18442	18500	40	18157	18500																						

- What does PrivacySuppressed mean?
- Why are there so many of them?
- Some variables are 100% “PrivacySuppressed”!

Technical Documentation: College Scorecard Data by Field of Study

Contents

Introduction	1
Data Methodology	1
<i>Defining field of study</i>	2
<i>Number of credentials conferred</i>	2
<i>Post-completion earnings</i>	3
<i>Cumulative loan debt</i>	4
<i>Repayment rate</i>	6
<i>Privacy protection</i>	7
Data Sources and Limitations	9
<i>Federally aided students</i>	9
<i>Earnings timepoints</i>	10
<i>Economic conditions</i>	11
<i>Aggregation of data</i>	13
<i>Monthly estimates</i>	13
<i>Nonemployment</i>	14
<i>Matching fields of study reported to different data systems</i>	14

- Enables us to understand what “PrivacySuppressed” means
- Enables us to understand the data generating process

Conclusion

- We can combine the data provided by the College Scorecard with data from users Facebook/Gmail profiles with their tastes and preferences to provide recommendations for programs of study
- The best RS is the one that drives the most user engagement
- We will still provide search capabilities, but the primary set of recommendations ought to come from the RS (just like Netflix)
- The MVP we develop is an iterable that improves with time:
 - The more users engage, the more data we collect
 - The more data we collect, the better the recommendations
 - The better the recommendations, the more users engage 🧐