In-State Workforce Retention of Utah's Postsecondary Graduates

Cory Stahle & Tyler Cain

February 2020

Abstract

The state of Utah values postsecondary graduates who remain in the state as they become contributors to a well-educated workforce. However, little is known about the factors that influence their decision to stay in Utah. This study aims to identify the characteristics that impact workforce retention following graduation. To accomplish this, data were obtained from the Utah Data Research Center (UDRC), which incorporates data from the Utah Department of Workforce Services, the Utah System of Higher Education (USHE), and the Utah System of Technical Colleges (UTech). The study used two cohorts: students who graduated from a USHE institution in 2013 and those who completed an award through UTech in 2013. Analysis included a retrospective look at retention by various characteristics of the graduates including: geographic origin, gender, age, race/ethnicity, field of study, award level, graduating institution, and financial aid.

Our analysis suggests that USHE and UTech students who originated from Utah or worked before graduation had the highest likelihood of remaining in the state after graduation. This study also predicted the retention rates of the 2016 cohort using gradient boosted decision trees and artificial neural networks. For 2021, it is predicted that 70.76% of USHE's graduates and 72.63% of UTech's graduates will be retained in Utah's workforce.

1 | Introduction

In-state retention of postsecondary graduates maintains and develops a well-educated workforce (Yazback, 2005). To maintain a well-educated workforce, many states have offered tuition support and loan repayment incentives similar to that created by Utah. In 2018, the state of Utah passed Senate Bill 104, which created loan repayment incentives to encourage postsecondary students with in-demand skills to remain in Utah after graduation. As a result, many studies have evaluated these programs and found a positive relationship between state administered tuition assistance and workforce retention (Harrington, Muñoz, Curs, & Ehlert, 2016; Leguizamon, Hammond, 2015). In contrast, literature related to retention predictors other than state administered aid is limited or narrow in scope.

In previous research, factors influencing workforce retention or migration of graduates are generally grouped into one of three categories: economic, noneconomic, and individual characteristics (Ishitani, 2011). Economic variables impact workforce retention as migrating individuals tend to move to states with stronger economies — measured by higher employment growth, lower unemployment, higher pay, and lower housing costs (Kodrzycki, 2001). Students also make post-graduation employment decisions based on noneconomic factors, such as family ties, geography, or available amenities (Groen 2004; Ishitani 2011). Individual characteristics and demographics have also been used to predict workforce retention with reasonable accuracy (74%) for students completing graduate medical education in Michigan (Koehler, Goodfellow, Davis, Spybrook, VanSchagen, & Schuh 2017).

This research builds upon the existing literature by specifically using individual characteristics to explore and predict post-graduation retention in Utah's workforce. Previous to this study, little was known about the students who remain in Utah after graduation and the factors that influence that decision. A report by the Utah Medical Education Council provides an annual measure of graduate workforce retention in Utah — but only for physicians. Additionally, the Utah System of Higher Education (USHE) provides data related to employment after graduation on their website but does not address the underlying influences of employment.

In order to examine graduate workforce retention in Utah and the factors that impact it, this study performed a historical analysis of retention partitioned by various student characteristics and used machine learning to predict future retention rates. The following questions are addressed in the study: 1) How many USHE and Utah System of Technical Colleges (UTech) graduates appear in Utah's workforce following degree or certificate completion? 2) Historically, how have rates of workforce retention varied across demographics or other factors? 3) What factors have significant influence on the decision to stay in Utah following graduation? 4) To what degree can individual characteristics predict whether graduates will stay in the state following completion? Gaining a greater understanding of the motivating factors behind post-graduation retention will enhance the ability of the state of Utah, the Utah Department of Workforce Services (Workforce Services), USHE, and UTech to engage in informed policy making to support Utah's recent graduates.

2 | Methods

2.1 | Workforce Retention Definition

Workforce retention, for the purpose of this study, was defined by workforce outcomes in Utah for each student. For one-year retention, a graduate was considered retained if any wages were reported to Workforce Services' unemployment insurance (UI) system in the calendar year following their graduation. An identical approach was used in calculating five-year retention.

2.2 | Study Group

A cohort study was used to understand the impact of demographics and education on workforce retention over time. Calendar year 2013 was selected as a base, as it allowed for one year of data prior to graduation and five years following. Graduates were included in the cohort if they graduated any time in 2013, and if they obtained a terminal award — defined as not receiving any additional awards following 2013. The analysis included all levels of awards offered through USHE institutions: certificates, associate, bachelor's, master's, and doctoral degrees. The UTech analysis included three certificate levels, each with a corresponding requisite number of necessary membership hours: certificates requiring less than 900 hours of education, those requiring between 900 and 1,800 hours, and certificates requiring more than 1,800 hours of training.

Additionally, only graduates born after 1973 were included in the cohort. Exclusion of older graduates allowed for a focus on graduates who were less likely to have established careers. Young graduates such as those who enrolled in an UTech postsecondary program while still in high school were included in the cohort. In total, 22,073 unique graduates were included in the final USHE cohort, and 4,833 unique graduates were included from the UTech cohort.

2.3 | Data

This study leveraged data from several sources integrated through the Utah Data Research Center (UDRC). Wage, employer, and industry of employer for graduates were acquired through the UI system — also known as wage records. The wage records were matched with administrative information from USHE and UTech. Inclusion of USHE and UTech data added information on completed awards, demographics (geographic origin, gender, race/ethnicity, and age) and financial aid or loans received.

2.4 | Data Transformations and Management

2.4.1 | Geographic Origin

More than 1,000 USHE graduates in the cohort were coded with an unknown geographic origin. Students with an unknown geographic origin, but had a Utah high school graduation assigned to their record, were assigned an in-state origin. After imputation, 177 graduates (0.8% of the total cohort) had an unknown geographic origin and were placed in the "unknown" category. Among the UTech cohort, the data had just three individuals with an unknown geographic origin. In total, four categories were included for origin: in-state, out-of-state, international, and unknown.

An additional variable indicating residency in a rural county in Utah prior to enrollment was derived from the data. Graduates who originated in Salt Lake, Utah, Davis, Weber, Cache, or Washington counties were categorized as urban residents. Graduates from the 23 remaining counties in Utah were identified as rural. These were selected to match the reporting of labor market information provided by Workforce Services. Graduates with out-of-state, international, and unknown geographic origins were assigned a value of "missing" since rural status could not be determined. Determination of urban or rural status was made for 67% (14,902 in-state origin graduates) of the USHE cohort and 96% (4,620 graduates in-state origin graduates) of the UTech cohort. UTech's service boundaries are mandated by law, which affects who it ultimately serves in the state. This may explain the difference of urban versus rural students who enroll in either system.

2.4.2 | Financial aid

To prepare USHE data for analysis, several aid and loan categories were aggregated. Money borrowed from the federal government or another source were summed into one loan category. Federal government aid, state/local government aid, institutional aid, and other known aid sources rolled up into a measure of aid received. Graduates were assigned a dummy variable (a binary representation of categorical data), indicating whether they accepted aid or loans of any nominal amount in the calendar year prior to graduation. Graduates were also assigned a dummy variable of participation in a work study program prior to award completion.

For UTech students, they may receive federal government aid through Pell grants, Bureau of Indian Affairs grants, or state government aid through Workforce Services' Workforce Innovation and Opportunity Act (WIOA) program. These types of aid that were awarded in the calendar year prior to graduation were summed to measure financial aid received for the UTech cohort. Students receiving financial aid through institutional scholarships, county scholarships, and other statutorily created and private scholarships are not reflected in this data. With respect to loans, none of the UTech institutions accept federal student loans.

2.4.3 | Field of Completion

Classification of Instructional Programs (CIP) codes are provided for graduates. This hierarchical coding system is used to indicate the academic field of study or completion. Codes associated with award completion, rather than intended award, rolled up to the two-digit level were used. In cases where graduates obtained multiple awards at the same time, a 0/1 flag for multiple award attainment was added and a CIP code was selected at random. Only one code was selected to prevent duplicate entries and bias across other demographic features. There were 414 USHE students and 19 UTech students who were flagged with multiple awards.

2.4.4 | Work Experience and Industry

In addition to wage information, UI records provide details regarding past employers and industries for graduates in Utah. Industry codes are reported using the North American Industry Classification System (NAICS). In cases where graduates held jobs with multiple employers in a calendar year, the employer responsible for paying the highest sum of wages in the year was designated as the primary employer for each individual. The NAICS used in the analysis represents the two-digit NAICS of each graduate's primary employer. Graduates without an identifiable employer were categorized as not appearing in the wage record. Wage records were also used to derive an additional variable for the analysis. For graduates in the wage record prior to graduation, employer identification numbers were used to generate an indicator of whether they remained with the same employer after award completion.

2.5 | Data Analysis

Two analyses per cohort were performed: historical workforce retention analysis and predictors of in-state retention of graduates. Analyses for USHE graduates were performed in Python (v 3.7) and leveraged several libraries including: Scikit-learn (v0.19.2), Pandas (v0.23.4), NumPy (v1.15.1), Statsmodels (v0.9.0), and Matplotlib (v2.2.3). The R software environment 3.6.0 (R Core Team, 2019), along with the Dplyr (v0.8.1, Wickham, 2019), Caret (v6.0.84, Kuhn, 2019), Tidyverse (v1.2.1, Wickham, 2017), RandomForest (v4.6.14, Liaw, 2018), Kernlab (v0.9.27, Karatzoglou, 2018), and E1071 (v1.7.2, Meyer, 2019) libraries, provided a platform for assessing UTech in-state retention of graduates.

2.5.1 | Workforce Retention Analysis

Rates of one-year and five-year workforce retention were calculated for the cohort and across several characteristics including: geographic origin, demographics (gender, age, race/ethnicity), award level, graduating institution, field of study (award CIP code), employment, and financial variables. Rate of workforce retention was calculated as the sum of graduates retained divided by total graduates in the cohort. One-tailed z-test for two proportions testing was performed to identify significant differences between retention rates by characteristics and the average rate across the cohort. The same testing was also performed to identify significant differences in rates within the characteristics of rural/urban status, gender, and age.

2.5.2 | Predicting In-State Retention of Graduates

In addition to the historical workforce retention analysis, the study utilized machine learning to predict future workforce retention rates. In order to prepare data for machine learning, dummy variables were created for all categorical variables, including: award level, institution name, geographic origin, CIP code, race/ethnicity, and NAICS code prior to graduation. Additionally, age of USHE graduates was scaled using Scikit-learn's StandardScaler class, while the UTech cohort used a binary classification (younger than 29 years old, and 29 years old or older).

Variable selection was performed to address data redundancy and reduce model training time using random forest feature importance. A random forest model was constructed using 125 decision trees, with each tree selecting the variables that maximized model accuracy in predicting workforce retention. To identify important features, variables were scored based on the mean reduction in impurity (error) across all nodes of the decision tree. Influential variables received higher scores than features with less impact. Importance scores were multiplied by 100 to improve readability. The result was the selection of 17 features for use in predicting one-year retention, and 16 features for five-year models for the USHE cohort. For the UTech cohort, the random forest model posited 30 features to predict one-year retention and 33 features for five-year retention.

Using the condensed datasets, split into a training set (75%) and test data (25%), eight classification models were constructed — four for each target (one- and five-year retention). In each model, a dummy variable for workforce retention was used as the target. Machine learning algorithms used included: gradient boosted decision trees, kernel support vector machines (SVM), k-nearest neighbors, and a multi-layer perceptron neural network. Each technique was selected for its ability to model and predict binary outcomes. Model parameters were tuned using grid search optimization (Appendix A). Each model was evaluated using k-Fold Cross-Validation with k equal to 10. Confusion matrices were created for the models. The models were trained using 5,519 unique graduates from USHE and 1,134 unique graduates from UTech.

3 | RESULTS

3.1 | Utah System of Higher Education

3.1.1 | Workforce Retention

Of the 22,073 graduates in the 2013 cohort, 15,415 were observed in the wage records (retained) one year following graduation (average cohort retention of 69.84%). Five years after completion, the rate of retention measured at 58.93%. Workforce retention varied by several graduate characteristics, particularly by employment history (Table 1).

3.1.2 | Geographic Origin

Graduates originating from any county in Utah were most likely to stay one year after graduation (77.25%) compared to out-of-state and international origins. Graduates of in-state origin also had the highest rate of workforce retention after five years at 66.97%. After five years, each geographic origin category had lower retention rates compared to one year (Fig. 1).

Graduates with out-of-state, international, or unknown geographic origins had workforce retention rates statistically significantly lower than the cohort average. In contrast, graduates with in-state origins had rates significantly above the cohort average — regardless of rural or urban origin. There was also a significant difference in one-year and five-year retention between graduates from rural counties and those from urban areas (Table 2).

Table 1: Count of USHE graduates in 2013 cohort, workforce retention rate, and Z-test for two proportions results by characteristic (Asterisks indicate significant difference from cohort average. Significance: *** = p < 0.001, ** = p < 0.01, * = p < 0.05).

	n	1-Year Retention		5-Year Retention	
		% Retained	z	% Retained	z
COHORT	22,073	69.84%	–	58.93%	–
ORIGIN					
In-State	15,738	77.25%***	15.78	66.97%***	15.71
Rural	2,297	78.03%***	5.19	65.65%***	6.16
Urban	12,605	75.14%***	16.28	67.69%***	16.00
Out-of-State	5,101	56.79%***	-18.08	44.44%***	-18.92
International	1,057	24.31%***	-31.04	13.25%***	-29.36
Unknown	177	58.19%***	-3.40	34.46%***	-6.61
EMPLOYMENT					
Worked in 2012	15,599	82.11%***	27.06	68.32%***	18.58
Did not work in 2012	6,474	40.27%***	-43.36	36.30%***	-32.12
DEMOGRAPHICS					
Sex					
Male	11,209	69.62%	-0.40	61.03%***	3.69
Female	10,864	70.06%	0.41	56.77%***	-3.75
Age					
Age: Under 29	15,482	69.27%	-1.18	57.33%***	-3.10
Age: 29 to 40	6,591	71.17%*	2.08	62.69%***	5.46
Ethnicity					
American Indian/Alaskan Native	160	56.88%***	-3.56	53.13%	-1.49
Asian	540	64.63%**	-2.60	52.04%***	-3.22
Black/African American	248	60.08%***	-3.33	47.58%***	-3.61
Hispanic/Latino	1,201	66.78%*	-2.25	54.29%***	-3.18
Multiple	239	70.71%	0.29	54.39%	-1.42
Native Hawaiian/Pacific Islander	123	67.48%	-0.57	63.41%	1.01
Non-Resident Alien	838	22.67%***	-28.75	13.01%***	-26.38
Unspecified	1,116	59.14%***	-7.56	53.49%***	-3.60
White	17,608	73.39%***	7.78	62.23%***	6.68

Table 1 (cont'd): Count of USHE graduates in 2013 cohort, workforce retention rate, and Z-test for two proportions results by characteristic (Asterisks indicate significant difference from cohort average. Significance: *** = p < 0.001, ** = p < 0.01, * = p < 0.05).

		1-Year Retention		5-Year Retention	
	n	% Retained	z	% Retained	z
AWARD LEVEL					
Associate's Degree	5,554	71.01%*	1.71	66.71%***	10.60
Bachelor's Degree	12,163	70.55%	1.38	56.17%***	-4.95
Certificate (<900h)	523	71.70%	0.92	63.29%*	2.00
Certificate (>=900h, <1800h)	182	76.92%*	2.08	68.13%**	2.51
Doctorate - Professional Practice	348	60.34%***	-3.82	56.90%	-0.77
Doctorate - Research	342	36.84%***	-13.13	21.93%***	-13.78
Master's Degree	2,709	68.51%	-1.42	58.73%	-0.20
Post-Baccalaureate Certificate	203	72.41%	0.80	57.64%	-0.37
Post-Master's Certificate	49	73.47%	0.55	71.43%*	1.78
INSTITUTION					
Salt Lake Community College	2,341	79.37%***	9.64	71.29%***	11.62
Weber State University	2,871	75.79%***	6.58	67.02%***	8.31
Dixie State University	1,090	73.85%**	2.83	65.59%***	4.37
Utah Valley University	3,637	71.26%*	1.75	60.46%*	1.73
University of Utah	6,177	70.10%	0.40	56.67%***	-3.18
Southern Utah University	1,196	67.14%*	-1.98	56.43%*	-1.71
Snow College	483	65.42%*	-2.09	68.53%***	4.24
Utah State University	4,278	59.25%***	-13.59	46.61%***	-14.90
FINANCE					
Loans					
Accepted loans (2012)	7,975	73.96%***	6.94	61.64%***	4.23
Did not accept loans (2012)	14,098	67.51%***	-4.67	57.40%**	-2.89
Financial Aid					
Received financial aid (2012)	11,343	71.03%*	2.26	59.03%	0.18
Did not receive financial aid (2012)	10,730	68.57%*	-2.33	58.83%	-0.18

Table 2: Statistical significance between USHE rural and urban graduates. Significance: *** = p < 0.001, ** = p < 0.01, * = p < 0.05.

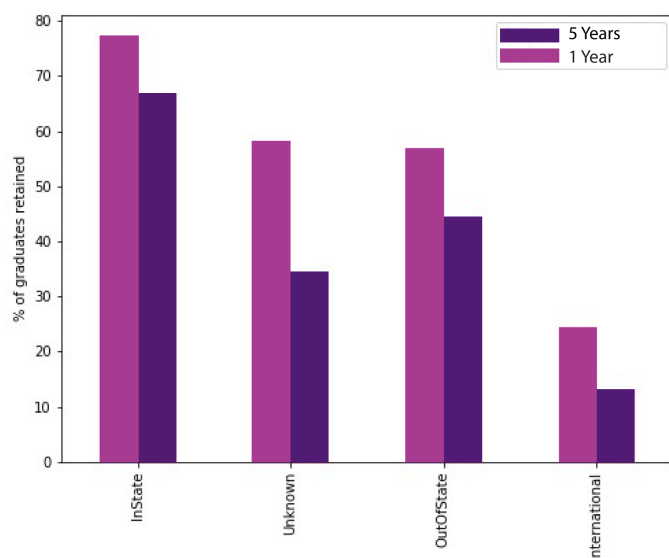|	n	% Retained 1 Year	% Retained 5 Years
Urban	12,605	78.03%	67.70%
Rural	2,297	75.14%	65.65%
z		-3.06***	-1.92*

Figure 1: Percent of USHE graduates retained by geographic origin — 1 and 5 years

3.1.3 | Employment History

There was a large difference in the workforce retention rates between individuals who worked before graduation and those who did not. Graduates who worked prior to completing their award were retained at a rate of 82.11% — significantly above the cohort mean (Table 1). Conversely, the retention rate for those who did not work (40.27%) fell significantly below average. A similar relationship and significance were also observed five years after graduation, as previously employed graduates were retained at a higher rate than those who did not work. Workforce retention rate differences remained large even after controlling for origin and level of degree obtained (Appendix B).

3.1.4 | Demographics

The study also analyzed workforce retention based on the self-reported gender of graduates. Rates of retention one year after graduation were nearly identical between males and females. Furthermore, one-year retention rates for males and females did not differ significantly from the cohort mean (Table 1) — or from each other. However, after five years, significant differences from the cohort mean had developed, with an above-average retention rate for males and below average retention rate for females. Retention rate changes for males and females also led to a statistically significant difference from each other, with a lower percentage of employment for females five years after graduation (Table 3).

When broken out by age, graduates ages 29-40 had significantly higher rates of workforce retention one year after completing school than individuals under the age of 29. The same relationship held true for five-year retention.

Workforce retention rates also varied by the race/ethnicity of graduates (Table 1). Individuals who self-identified as White had the highest rate of retention after one year, followed by those reporting multiple ethnicities, and Native Hawaiian or Pacific Islander. The three ethnic/racial identities or groups with the lowest workforce retention were Non-Resident Alien, American Indian or Alaskan Native, and individuals with an unspecified race/ethnicity. The rate of retention was lower for every race/ethnicity five years after award completion (Table 1).

Table 3: Statistical significance between USHE male and female graduates. Significance: *** = p < 0.001, ** = p < 0.01, * = p < 0.05.

	n	% Retained 1 Year	% Retained 5 Years
Male	11,209	69.62%	61.03%
Female	10,864	70.06%	56.77%
z		0.703	-4.517***

3.1.5 | Award Level

In general, certificate earners had the highest rates of workforce retention one year after graduation (Fig. 2). Recipients of associate degrees or long-term certificates had a retention rate significantly higher than the cohort average, while doctorate recipients were significantly lower. Graduates receiving a bachelor's degree or master's degree remained in the state at a rate not statistically different from the average. The five-year retention rates for recipients

of an associate degree, short-term certificate (< 900 hours), long-term certificate (>= 900 hours, < 1,800 hours), or post-master's certificate all measured significantly above average, while recipients of research-focused doctorate degrees or bachelor's degrees had lower than average workforce retention rates (Table 1).

Workforce retention of graduates varied based on the CIP code attached to the award completed. Communications technologies, precision production, and education led out with the highest rates of retention one year after graduation.
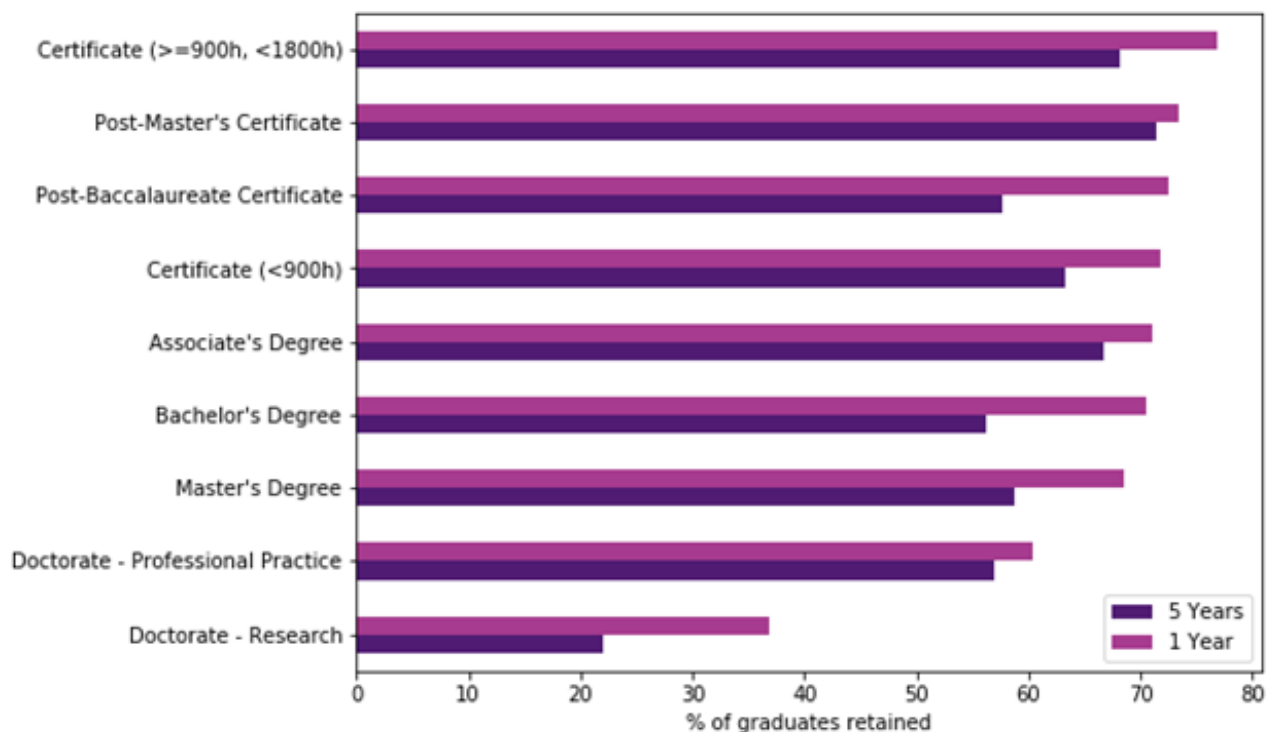

Figure 2: Percent of USHE graduates retained by award level— 1 and 5 years

Individuals completing awards in transportation and materials moving had the lowest rates of retention in both one year and five years following school completion.

While workforce retention rates varied by CIP code (Fig. 4), further analysis found interactions between CIP choice and geographic origin. For students of international origin, the majority (55.63%) were concentrated in social sciences, business management, or engineering awards. Furthermore, 44.52% of graduates with an out-of-state origin received an award in health professions and related clinical sciences, liberal arts/general studies, or business management. Unfortunately, many CIP codes have fewer than 10 graduates from outside of Utah. These potentially identifiable results cannot be published due to UDRC's disclosure protocols.

The institution attended had a significant correlation with workforce retention after graduation (Fig. 3). Overall, Salt Lake Community College had the highest one-year rate of retention at 79.37%, while the lowest rate of retention was observed in Utah State University graduates at 59.26%. Workforce retention rates were lower after five years in all but one institution — Snow College had a five-year retention rate above the one-year measure. When compared to the average rate of retention, Salt Lake Community College, Weber State University, Dixie State University, and Utah Valley University all had rates significantly higher than the mean. The University of Utah's workforce retention rate was not significantly different from the average. Three institutions — Southern Utah University, Utah State University, and Snow College — had retention rates below the average (Table 1).
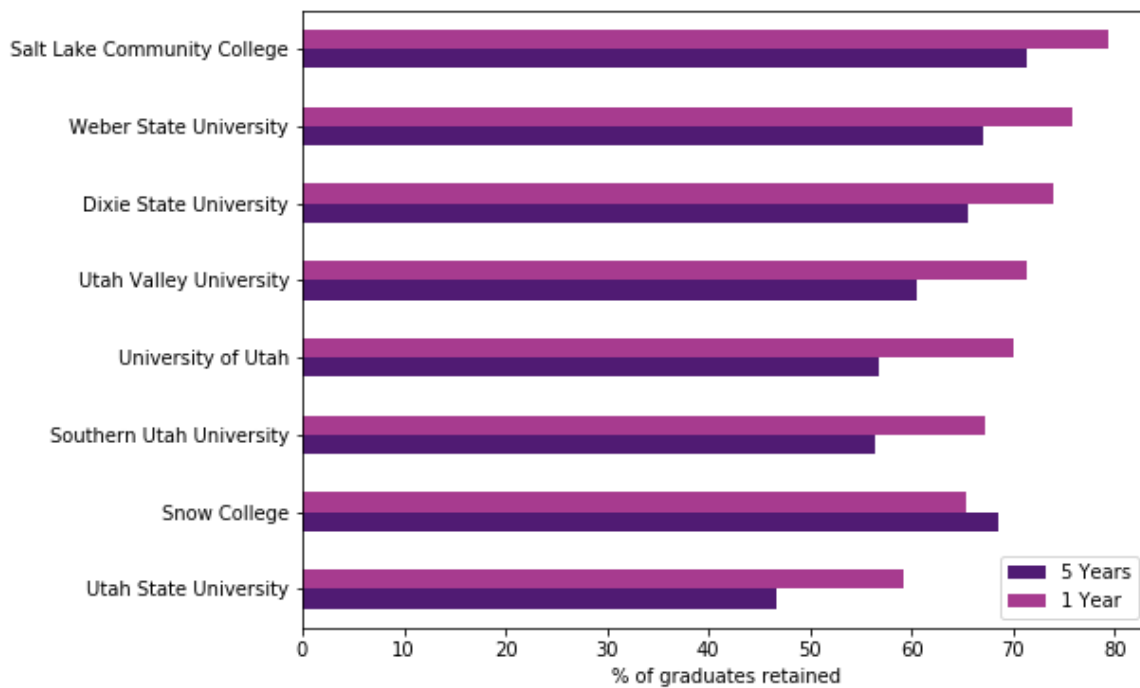
Figure 3: Percent of USHE graduates retained by institution.

3.1.8 | Financial Aid

Students who accepted loans or received financial aid the year prior to graduation were retained in the workforce at higher rates than those who did not. Workforce retention rates were statistically significant for those who accepted loans at both the one- and five-year marks. For those who received financial aid, only the one-year retention rate was statistically significant. Financial aid included government, state, local, institutional, and other known sources of aid.

3.1.9 | Predicting In-State Retention of USHE Graduates

In general, the graduate features that had the largest influence on workforce retention for USHE graduates — one and five years after award completion — were those related to employment and geographic origin (measures of origin, residency, rural, or high school attended). Of the 17 variables selected for the one-year model, nine were related to the geographic origin of graduates. Similarly, nine of the 16 variables used in five-year estimates reflected geographic origin (Table 4). In both models, the most influential predictor was whether a student worked before graduation or not. The importance score measures the magnitude of influence on retention and not whether that feature has a positive or negative impact.

The first model in the predictive modeling analysis was a gradient boosted decision tree. When applied to test data, the boosted decision tree was able to correctly predict one-year retention with an average 77.37% accuracy rate. Cross-validation found that the model was also stable (least deviation) and had a standard deviation of 0.71%. The second model, kernel SVM, had a slightly lower average accuracy (77.06%) than the boosted decision tree, but also had a lower standard deviation (0.69%). A k-Nearest Neighbors model was then applied to the data and yielded the most stable, but least accurate model in the analysis. The final model used was an artificial neural network (MLP). The neural network was able to predict cases of workforce retention or non-retention with 77.32% accuracy. Overall, each model was biased towards Type 1 errors — predictions that graduates would stay, but did not (Appendix D). Table 5 contains a summary of accuracy and deviation of models used to predict one-year retention.

A gradient boosted decision tree was selected for its accuracy and stability to predict one-year workforce retention for the 2016 cohort. In total, the 2016 cohort included 29,351 graduates. Of these, the model predicted that 23,084 — or 78.65% — would remain in Utah for a year following graduation. In reality, the 2016 cohort had 21,350 graduates — or 72.74% — remain in Utah a year following graduation. This actual workforce retention rate was 5.91% lower than the predicted rate.

The same four modeling techniques were used

to predict five-year retention. Overall, the mean accuracy decreased with each model in predicting five-year retention. However, the variation in accuracy between the models remained low. Overall, the kernel SVM had the highest average accuracy, but also the highest deviation. The artificial neural network measured second in ability to predict cases of workforce retention, and had the lowest standard deviation. The Gradient boosted decision tree and k-Nearest neighbors followed in accuracy, and had deviations similar to the other models (Table 6). Overall, each model was biased towards Type 1 errors (Appendix D).

Of these models, the artificial neural network was selected to predict five-year retention given its high accuracy and low deviation relative to the other models. Of the 29,351 graduates in the 2016 cohort, 20,768 (70.76%) were predicted to be retained in 2021 (Appendix E).

Table 4: Top five predictor variables of in-state workforce retention for USHE graduates from the one- and five-year modeling (from random forest feature importance; Importance = mean reduction in node impurity * 100).

One-year selected variables		Five-year selected variables	
Feature	Importance	Feature	Importance
Worked before graduation	33.944	Worked before graduation	22.335
Paid nonresident tuition at any time	13.303	Paid nonresident tuition at any time	13.233
In-state origin	6.565	In-state origin	9.546
Unknown rurality status	5.815	Attended Utah high school	7.642
International origin	5.391	Unknown rurality status	7.310

Table 5: Model performance for predicting one-year retention for USHE graduates. Model evaluated with 10-Fold cross-validation.

Model	Mean Accuracy	Standard Deviation (σ)
Gradient Boosted Decision Tree	77.37%	0.71%
Kernel SVM	77.06%	0.69%
k-Nearest Neighbors	76.31%	0.43%
Artifical Neural Network (MLP)	77.32%	0.78%

Table 6: Model performance for predicting five-year retention for USHE graduates. Model evaluated with 10-Fold cross-validation.

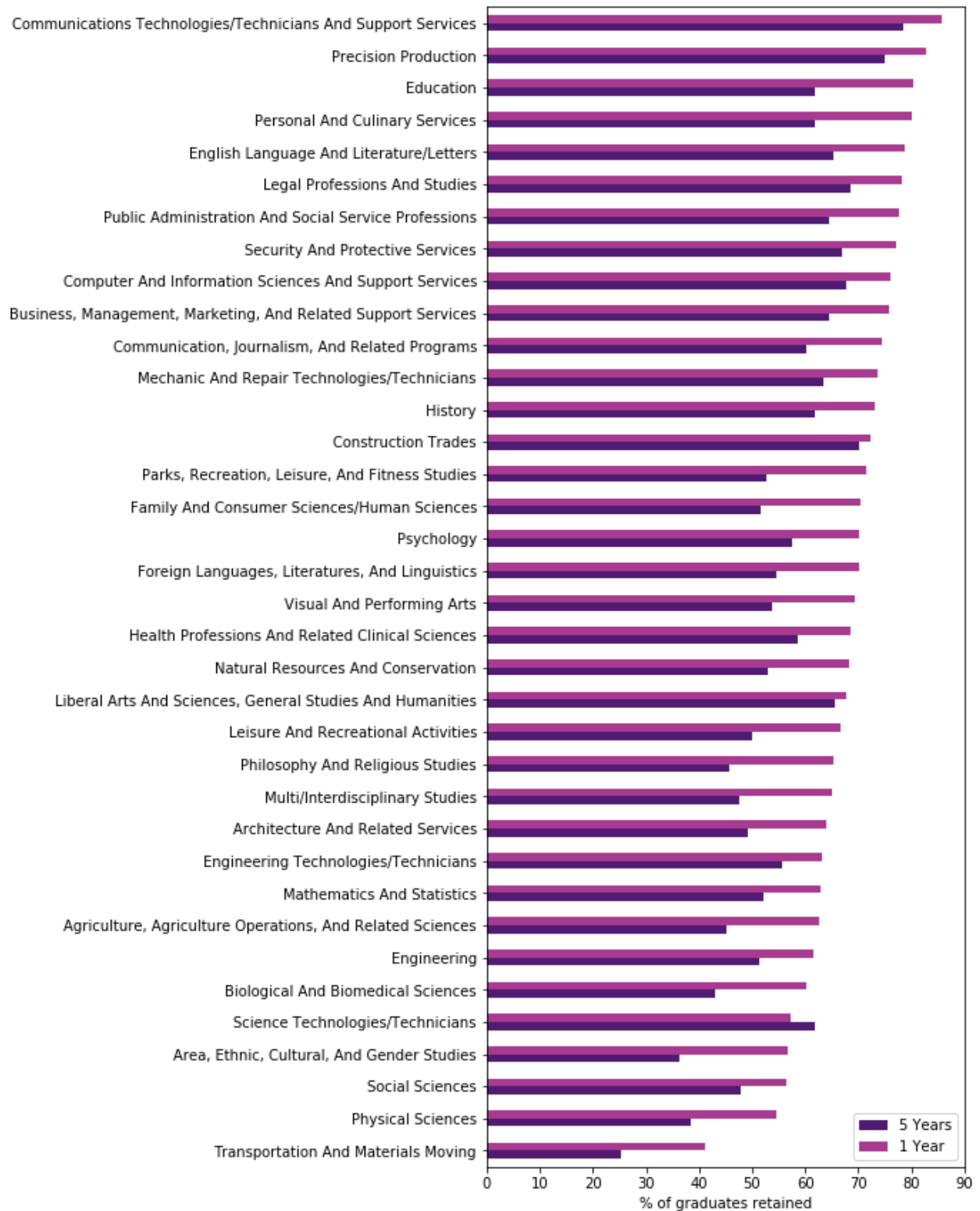| Model	Mean Accuracy	Standard Deviation (σ)
Gradient Boosted Decision Tree	69.36%	0.76%
Kernel SVM	69.54%	0.84%
k-Nearest Neighbors	67.90%	0.80%
Artifical Neural Network (MLP)	69.51%	0.64%

Figure 4: Percent of USHE graduates retained by CIP code — 1 and 5 years

3.2 | Utah System of Technical Colleges

3.2.1 | Workforce Retention

Of the 4,833 graduates in the 2013 UTech cohort, 80.01% of the individuals were identified in wage records one year after certificate completion. At the five-year mark, 72.95% from the cohort were retained. Previous employment, gender, and the type of certificate obtained proved to be the most salient factors when analyzing workforce retention (Table 7).

3.2.2 | Geographic Origin

The UTech cohort was mostly comprised of in-state students, with 95.59% UTech graduates who reported an in-state residence. Conversely, 4.26% indicated they were from outside the state of Utah. Lastly, seven graduates reported an "unknown" geographic origin. These individuals were not subject to subsequent analysis.

Graduates originating from any county in Utah were 30% more likely to remain in the state one year after certificate completion when compared to their out-of-state counterparts. At the five-year mark, 74.80% of in-state graduates were retained, as reflected in wage records; contrasted to 32.03% of their out-of-state peers (Fig. 5).

Significance testing supported the inference that graduates with an out-of-state origin had lower workforce retention rates. Graduates reporting an out-of-state origin demonstrated statistically significant lower workforce retention rates both one year and five years after certificate completion.

Within the in-state UTech graduate pool, 79.16% of students came from urban counties and 20.84% came from rural counties. Rural students had a higher workforce retention rate at the one-year mark, though were outpaced by their urban counterparts at five years. 794 rural graduates remained in the local economy one year after certificate completion, and 715 of those individuals remained five years after. Rural students had a statistically significant higher workforce retention rate at the one-year mark when compared to the overall mean of UTech graduates. At the five-year mark, rural students still had a higher workforce retention rate than the overall mean of UTech graduates, but it was not statistically significant.

3.2.3 | Employment History

This variable was the most decisive in predicting longitudinal workforce retention. UTech graduates who were employed one year prior to certificate completion were retained at 88.97%, which was above the cohort mean. This trend persisted at the five-year mark, as well; with 79.95% of previously employed graduates retained. These results were statistically significant at the highest level. For graduates who were not previously employed prior to certificate conferral, 60.67% were retained at the one-year mark, and steadily declined after five years, which was statistically significant (Table 7). Workforce retention rates remained higher for those who worked compared to those who did not work when controlling for geographic origin and level of degree obtained (Appendix F).

3.2.4 | Demographics

The UTech graduate pool consisted of 43.12% males and 56.65% females. Despite the higher proportion of women in the cohort, workforce retention rates did not diverge from the cohort mean both one and five years after certificate completion (Table 7). However, retention rates varied when men and women were compared to each other. Women experienced higher retention rates at one year, and men experienced higher retention rates at five years (Table 8).

When examined by the context of age, UTech graduates under 29 years of age and graduates from 29-40 years old did not have statistically significant workforce retention rates at either the one- or five-year marks post-completion. Additionally, these groups were not found to be meaningfully different when compared to each other at both the one- and five-year time increments. Instead, when assessing both age groups at both time periods, the retention rates hovered around the cohort mean.

Workforce retention rates also varied across time for race and ethnicity. Individuals who self-identified as Asian or Black/African American experienced retention rates below the cohort mean at both timing increments, with only some of the relationships being statistically significant (Table 7). Graduates who self-identified as White had the highest rates of retention both one and five years after certificate completion, both of which exceeded the cohort mean. Black/African American students had the lowest rate of retention one year after certificate completion, followed by students who self-identified as Asian and graduates with an unspecified race.

Students of Pacific Islander descent experienced the largest decrease in retention from one to five years. White students were the only racial/ethnic group to have both a positive and statistically significant retention rate five years after certificate completion. The cumulative number of students who self-identified with respect to their ethnic/racial background exceeded the total number of students in the cohort, so it is possible that students of multiple races selected more than one race in their student documentation and were counted twice.

Table 7: Count of UTech graduates in 2013 cohort, workforce retention rate, and Z-test for two proportions results by characteristic (Asterisks indicate significant difference from cohort average. Significance: *** = p < 0.001, ** = p < 0.01, * = p < 0.05).

	n	1-Year Retention		5-Year Retention	
		% Retained	z	% Retained	z
COHORT	4,833	80.01%	–	72.95%	–
ORIGIN					
In-State	4,620	81.29%*	1.56	74.81%*	2.02
Rural	963	82.55%*	1.77	74.35%	0.85
Urban	3,657	80.97%	1.07	74.92%*	2.01
Out-of-State	206	50.97%***	-9.90	32.03%***	-12.63
Unknown	7	85.71%	0.00	85.71%	0.33
EMPLOYMENT					
Worked in 2012	3,302	88.97%***	10.70	79.95%***	7.21
Did not work in 2012	1,531	60.67%***	-15.26	57.87%***	-11.12
DEMOGRAPHICS					
Sex					
Male	2,084	78.84%	-1.17	74.71%	1.49
Female	2,738	81.08%	1.09	71.69%	-1.15
Unspecified	11	54.54%*	-1.73	54.54%	-1.03
Age					
Age: Under 29	3,789	79.49%	-0.57	72.81%	-0.12
Age: 29 to 40	909	81.41%	0.92	73.16%	0.08
Ethnicity					
American Indian/Alaskan Native	61	75.40%	-0.73	62.22%*	-1.71
Asian	78	69.23%**	-2.12	66.67%	-1.11
Black/African American	82	62.19%***	-3.84	58.53%**	-2.78
Hispanic/Latino	383	77.80%	-0.97	68.66%*	-1.75
Native Hawaiian/Pacific Islander	48	77.08%	-0.32	56.25%**	-2.42
Unspecified	450	76.00%*	-1.96	68.88%*	-1.79
White	3,806	81.31%	1.49	74.61%*	1.71
AWARD LEVEL					
Certificate (<900h)	3588	71.98%***	-8.57	73.16%	0.18
Certificate (>=900h, <1800h)	1044	84.48%***	3.28	73.94%	0.62
Certificate (>=1800)	201	75.62%	-1.34	64.17%	-2.65

Table 7 (cont'd): Count of UTech graduates in 2013 cohort, workforce retention rate, and Z-test for two proportions results by characteristic (Asterisks indicate significant difference from cohort average. Significance: *** = p < .001, ** = p < .01, * = p < .05).

	n	1-Year Retention % Retained	z	5-Year Retention % Retained	z
INSTITUTION					
Bridgerland Technical College	777	79.15%	-0.51	70.91%	-1.14
Davis Technical College	1,063	75.25%***	-3.41	68.203%***	-3.09
Dixie Technical College	127	79.52%	-0.02	76.37%	0.76
Mountainland Technical College	1,325	81.59%	1.24	78.64%***	4.15
Ogden-Weber Technical College	861	84.55%**	3.06	77.46%**	2.72
Southwest Technical College	139	81.29%	0.27	67.62%	-1.29
Tooele Technical College	79	83.55%	0.64	79.74%	1.22
Uintah Basin Technical College	462	78.57%	-0.68	62.12%***	-4.90
FINANCE					
Financial Aid					
Received financial aid (2012)	327	85.32%*	2.30	65.14%**	2.97
Did not receive financial aid (2012)	4,511	80.67%	0.81	73.59%	0.758

3.2.5 | Award Level

Students who obtained certificates requiring fewer than 900 hours of education had the lowest workforce retention rate one year after certificate completion, which demonstrated the highest level of statistical significance. Students who obtained certificates requiring beyond 900, but less than 1,800 hours of education, had the highest retention rate after one year. Graduates with certificates requiring beyond 1,800 hours did not have a statistically significant difference in retention after one year, though the group was slightly below the cohort average. None of the various certificate category groups significantly deviated from the cohort mean five years post completion (Fig. 6).

Table 8: Statistical significance of UTech male and female retention rates. Significance: *** = p < 0.001, ** = p < 0.01, * = p < 0.05.

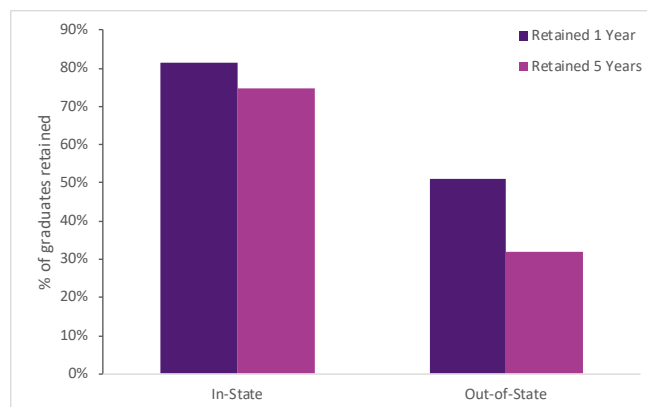|	n	% Retained 1 Year	% Retained 5 Years
Male	2,084	78.84%	74.71%
Female	2,738	81.08%	71.69%
z		-1.977*	-2.305*

Figure 5: Percent of UTech graduates retained by geographic origin — 1 and 5 years.

3.2.6 | Institution

The institution a student attended also had a significant correlation with workforce retention in certain cases. Specifically, individuals who graduated from Ogden-Weber Technical College had the highest retention rate one year after certificate completion (Fig. 7). This positive relationship between retention and graduating from Ogden-Weber Technical College also carried through to the five year mark. After five years, positive retention rates were associated with graduates from Dixie Technical College, Mountainland Technical College, Ogden-Weber Technical College, and Tooele
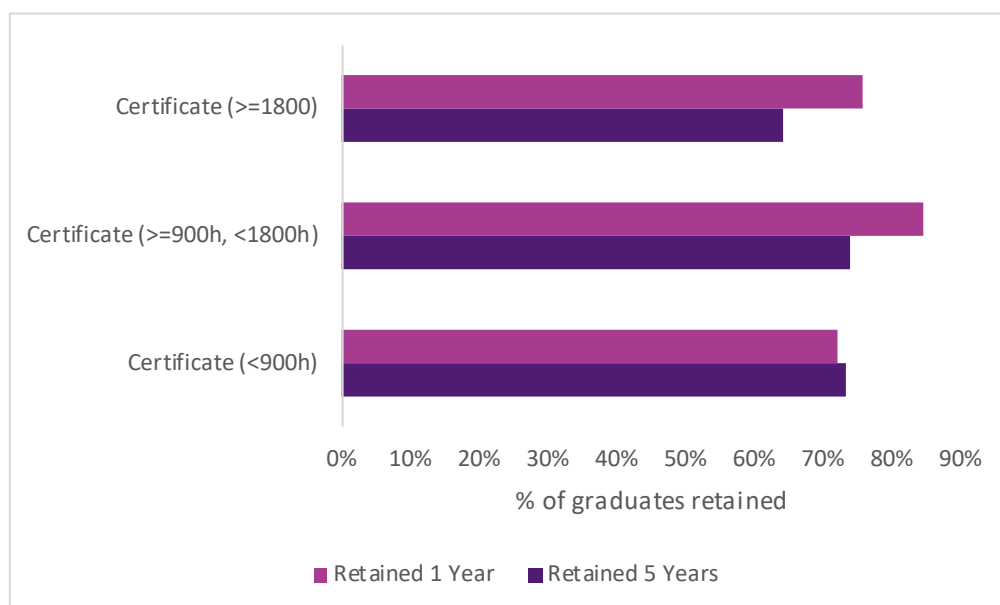
Figure 6: Percent of UTech graduates retained by level of award — 1 and 5 years.

Technical College. However, only Mountainland Technical College and Ogden-Weber Technical College retention rates proved to be both positive and statistically significant from the mean after five years. Graduates from Davis Technical College and Uintah Basin Technical College both had lower retention rates significantly different from the cohort average after five years (Table 7).

3.2.7 | Field of Study

Workforce retention of UTech graduates varied based on the CIP code attached to the completed certificate. Health professions and related clinical sciences certificate graduates accounted for the largest enrollment numbers, and also displayed proportional minimal attrition when compared to other CIP codes. Additionally, students who completed certificates related to precision

production, construction trades, and engineering technology maintained steady retention rates when evaluated at both the one- and five-year increments (Fig. 8).

3.2.8 | Financial Aid

At the one-year mark, students who received financial aid the year prior to graduation were retained in the workforce at a higher rate than those who did not. However, those who did not receive financial aid were retained at a higher rate at the five-year mark. The relationship with workforce retention was significant only for those who received financial aid. These results reflect only federal and state government aid, and does not include institutional, county, or private aid. Loans are not included since UTech does not accept federal student loans for its programs.

Table 9: Top five predictor variables of in-state workforce retention for UTech graduates from the one- and five-year modeling (from random forest feature importance; Importance = mean reduction in node impurity * 100).

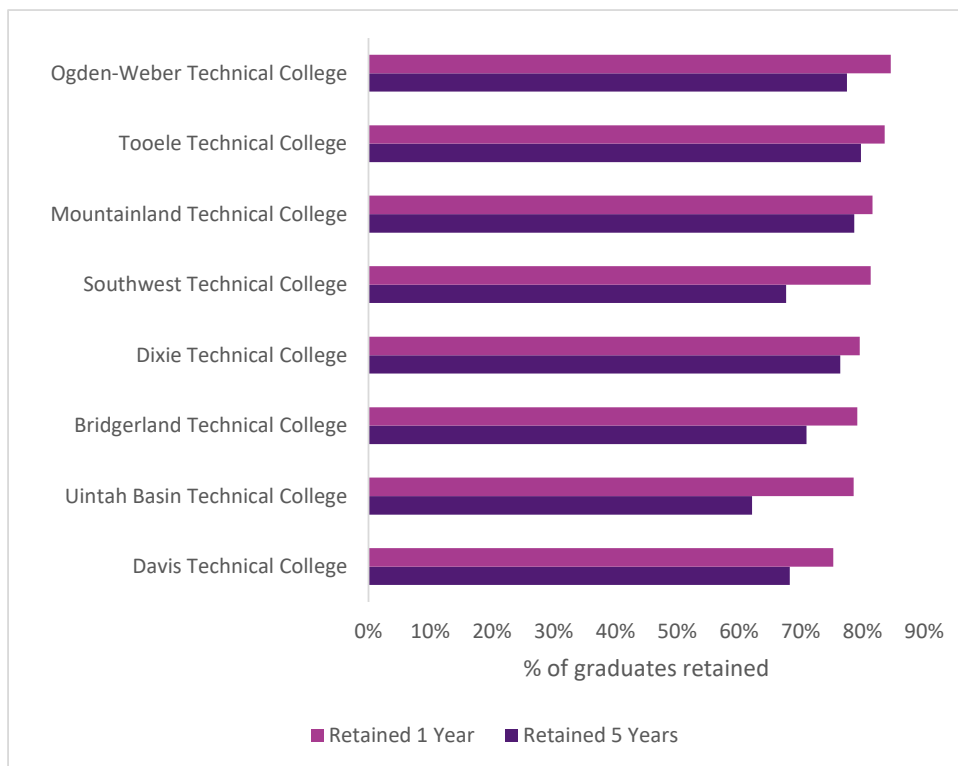| One-year selected variables		5-year selected variables	
Feature	**Importance**	**Feature**	**Importance**
Worked Before Completion	48.457	Worked Before Completion	30.953
Gender	10.313	Age	13.796
Age	9.762	Gender	12.207
Urban Utah Origin	8.281	Certificate Type	10.452
Certificate Type	7.215	Urban Utah Origin	9.705

Figure 7: Percent of UTech graduates retained by institution — 1 and 5 years.

3.2.9 | Predicting In-State Retention of UTech Graduates

Among all of the variables assessed, previous employment status prior to certificate completion, gender, and graduate age had the largest influences on both one- and five-year workforce retention rates for UTech graduates (Table 9). To model the one-year retention rates, a total of 30 variables were selected according to their influence with 10 of the variables related to the field of study. The five-year model incorporated 33 variables based on influence, 11 reflecting selected field of study. In both models, the most influential predictor was whether a graduate worked before award completion or not. The importance score measures the magnitude of influence on workforce retention and not whether that feature has a positive or negative impact.

This process of the analysis mirrored that of the USHE cohort, with a gradient boosted decision tree being the first model utilized to predict workforce retention. After training, the boosted decision tree accurately predicted 81.83% of the outcomes with a 0.40% standard deviation when applied to the test data. The gradient boosted decision tree represented both the most accurate and the second most stable (less deviation) model among the others. The second model employed, the kernel SVM, had a 79.79% average prediction accuracy with a standard deviation of 0.41%. Among the four predictive modeling techniques employed, the kernel SVM accounted for the lowest average accuracy. The k-Nearest Neighbors model yielded an average

accuracy of predicting 80.05% with a standard deviation of 0.41%. Lastly, and closely following the gradient boosted decision tree in terms of accuracy, the artificial neural network (MLP) model predicted 81.57% of the outcomes with a standard deviation of 0.39%. The artificial neural network proved to be the most stable among all of the modeling techniques. Overall, each model was biased towards Type 1 errors — predictions that graduates would stay, but did not (Appendix H).

The artificial neural network model was selected for its high accuracy and being the most stable to predict one-year workforce retention for the 2016 cohort. In total, the 2016 cohort included 5,933 graduates who completed a certificate. Of these, the model predicted that 4,513 graduates — or 76.07% — would remain in Utah for a year after completion. In reality, the 2016 cohort had 4,855 — or 81.83% — remain in Utah a year following completion. The actual workforce retention rate was 5.76% higher than the predicted rate.

Average levels of accuracy dropped slightly among the same four modeling techniques when used to predict five-year retention. However, variation between the models remained low overall with nearly the same amount of deviation. The artificial neural network was the most accurate in predicting workforce retention, and also displayed the lowest standard deviation among the models. The gradient boosted decision tree followed close behind the artificial neural network in terms of accuracy and minimal deviation, and was the second-best

in predicting workforce retention (Table 11). The k-Nearest Neighbors and kernel SVM followed both in accuracy and in deviation, with SVM returning the highest standard deviation score by a slight amount among all of the models. Overall, each model was biased towards Type 1 errors (Appendix H).

The artificial neural network model was selected for having the most accurate and stable model to predict five-year retention for the 2016 cohort. It predicted that 4,309 graduates — or 72.63% — would be retained in 2021. For context, descriptive statistics related to the characteristics of the 2016 cohort can be found in Appendix I. For comparison, UTech's predicted rate differed by 2% from USHE's predicted rate for the 2016 cohort. Previously, there was a 14% difference between the two systems when comparing the actual workforce retention rates of the 2013 cohort. This predicted closing gap of 2% is due to the prediction of USHE's workforce retention rates increasing over time. The 2016 cohort of both systems may meet more requirements of the model's most important features more so than in 2013. For example, more USHE graduates in the 2016 cohort worked prior to enrollment in UTech than the 2013 cohort. Since work history was an

important feature, more USHE graduates may have been predicted to be retained since more of them met this requirement. In addition, the model did not account for economic conditions that may have changed over time with each cohort. If the model included economic conditions, the predicted workforce retention rate may have been more aligned with past rates for USHE or would have had additional covariates to explain why the model predicted a higher USHE workforce retention rate.

4 | Discussion

4.1 | Utah System of Higher Education

This study used individual characteristics to explore and predict post-graduation retention in Utah. Results found that nearly seven in 10 graduates obtain UI-covered employment within a year of graduation. After five years, about six in 10 graduates are still observed in Utah wage records. These findings suggest that graduates who are retained in the state one year after completion are likely to remain for five years. The least common workforce retention outcome for graduates was to
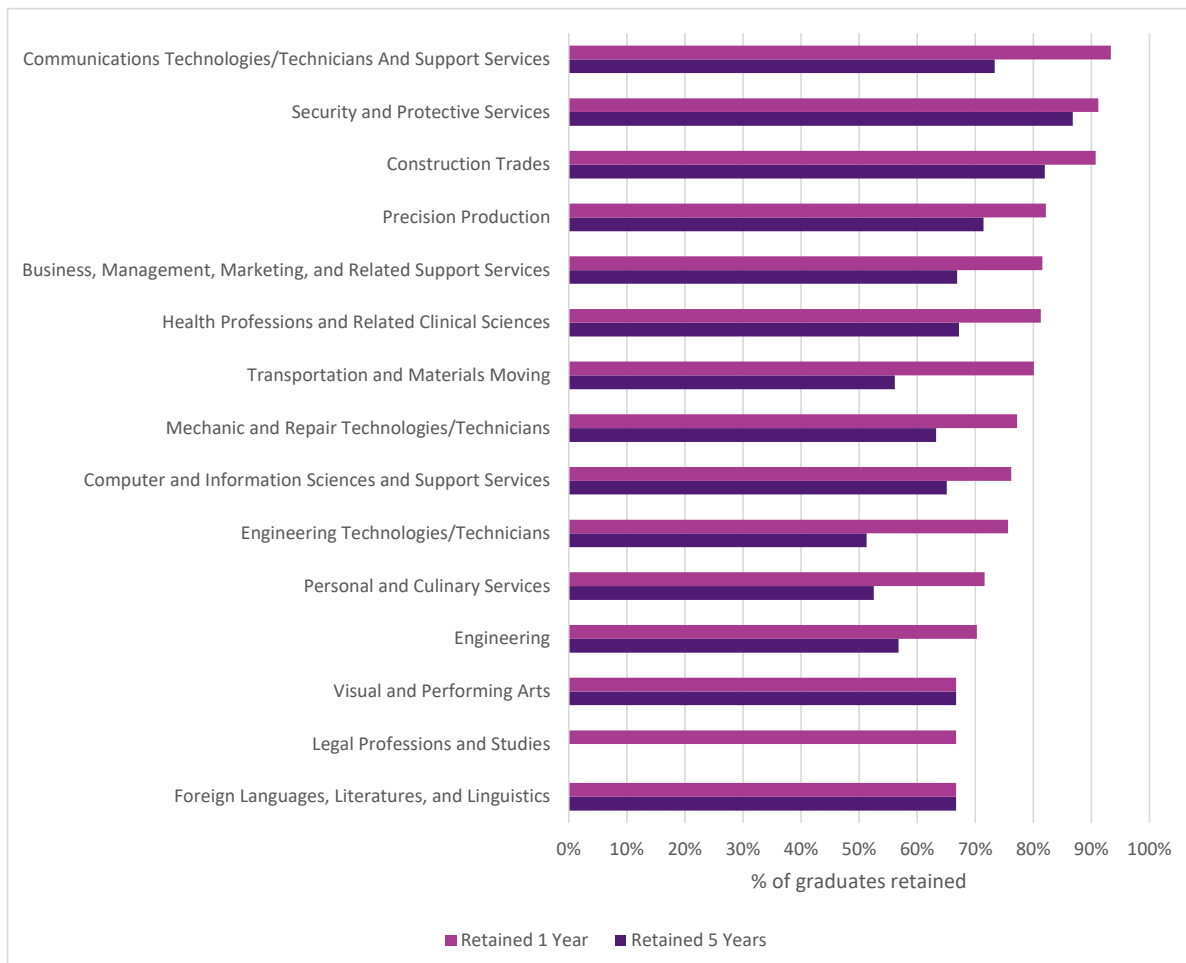

Figure 8: Percent of UTech graduates retained by CIP code (field of study) — 1 and 5 years.

Table 10: Model performance for predicting one-year workforce retention for UTech graduates.

Model	Mean Accuracy	Standard Deviation (σ)
Gradient Boosted Decision Tree	81.83%	0.40%
Kernel SVM	79.79%	0.41%
k-Nearest Neighbors	80.05%	0.41%
Artifical Neural Network (MLP)	81.57%	0.39%

Table 11: Model performance for predicting five-year workforce retention for UTech graduates.

Model	Mean Accuracy	Standard Deviation (σ)
Gradient Boosted Decision Tree	77.34%	0.42%
Kernel SVM	73.37%	0.46%
k-Nearest Neighbors	72.57%	0.45%
Artifical Neural Network (MLP)	74.51%	0.44%

not appear in the wage records after one year, but do so at five years following award completion.

Analysis also found that geography plays a significant role in determining workforce retention after award completion. Graduates who originate from Utah, regardless of rural or urban status, have retention rates significantly above the cohort average. In some cases, students may face legal constraints on remaining in the state following graduation — particularly non-resident aliens (international graduates). For those not facing such constraints, it is likely that qualitative factors such as family ties and geographic preference influence this decision, as suggested by Groen (2004) and Ishitani (2011). Qualitative influencers were not explored in this research; hence, future work is needed to better understand the impact of these factors.

In addition to geography, this study found that a graduate's employment history has a significant impact on workforce retention. Overall, working a year prior to graduation was the strongest predictor of whether a graduate would remain in Utah after award completion — even when controlling for geographic origin or award level. This may be due, in part, to the definition of retention used by the study, which measured employment outcomes. This definition may favor students with job experience and skills, which provide a labor market advantage after graduation. Conversely, those without work experience may face challenges in finding employment after graduation and may choose to leave the state or labor force as a result. Perhaps non-working students obtain credentials but lack the employment history or skills required to find a job. Exploring the connection between employment history, skills, and workforce retention warrants further study.

Student demographics, particularly gender and age, had little impact on one-year retention but became significant after five years. This suggests that students initially find employment after graduation and then leave in the years that follow. For females, this drop in employment over time may be the result of choosing to stay home to focus on raising children or pursue work alternatives like self-employment. For younger graduates (under 29 years) it is likely that they work immediately after graduation to build or begin accruing experience. Once they have that experience, new opportunities arise and their flexibility (fewer familial/ other responsibilities) allows them to pursue opportunities outside the state for employment. Further analysis into the interactions between gender and age is recommended. For race/ethnicity, workforce retention rates varied and were heavily informed by an interaction with geographical origin, as groups with lower retention rates had more out-of-state students (Appendix C).

Analysis also found a relationship between award level and retention. Results, particularly for five-year retention, suggest that completing a higher award level is correlated with a higher likelihood of leaving the state or UI-covered employment. While higher education generally offers greater career opportunities and may increase the likelihood of mobility, it is also important to consider other interactions. For instance, one finding indicated that individuals graduating with a research-focused doctorate were retained significantly below average. However, when broken out by geographic origin, two thirds of the graduates were not from Utah — a

group who is more likely to leave anyway. One of the reasons they leave the state is because standard academic practice discourages hiring graduates as faculty at institutions where they trained or sometimes in similar institutions. Similarly, the drop in workforce retention for bachelor's degree earners may reflect a legitimate movement of graduates after five years, but also reflects interactions with trends seen in other characteristics like age and gender. Future research could provide additional insight on this issue by exploring the interactions between graduate characteristics and economic opportunity.

This study also identified correlations between some graduating institutions and workforce retention. Results suggest that institutional workforce retention rates are a function of the composition of their enrolled students. Salt Lake Community College primarily educates local, job-focused candidates and has a high retention rate. Additionally, it only offers associate degrees and below, which are correlated with higher workforce retention in the analysis. In contrast, Utah State University has a lower workforce retention rate but serves students from around the world and the United States — particularly those from across the Utah-Idaho border. Graduates of agricultural programs offered through Utah State University may be excluded from the wage records due to aforementioned UI system reporting requirements. In other words, they may work in Utah but will not be counted as retained if they are not covered by Utah's UI program. Furthermore, Utah State University has several online degree offerings, which may impact their student composition and influence workforce retention.

Results also illustrated a difference in workforce retention by CIP code but suggested that it had little predictive power in determining retention. This does not mean that the field of study is not important; rather, it suggests that other factors underlie the choice of degree/award and subsequently influence retention. While several factors may explain choice of degree, the study only looked at geographic origin. Overall, it was observed that students from outside Utah tend to enroll in a small range of programs. Further analysis into the interactions between degree/award choice and individual characteristics is recommended.

Financial analysis indicated a significant tie between accepting loans and remaining in Utah after graduation; however, caution should be exercised in these findings. This study was limited and did not look at student finance in terms of eligibility. This creates problems in the findings since some graduate characteristics — such as citizenship — influence eligibility for these programs. If students

from outside Utah are more likely to leave after graduation, and those same graduates are not eligible to receive loans, the findings may be biased. Similarly, many aid programs have eligibility requirements that may preclude individuals who are less likely to be retained. Future research on financial assistance such as loan repayment programs, with a condition of eligibility, is recommended.

Finally, in line with Koehler et. al (2017), this study found that a combination of individual characteristics can be used to predict future workforce retention with reasonable accuracy. The largest loss in model accuracy stemmed from bias toward Type 1 errors. These errors overestimate the number of graduates who will be retained and predict a workforce retention rate that will likely be higher than reality. Despite this bias, the models illustrated significant correlations between characteristics and workforce retention. Future research may improve upon these findings by incorporating additional data and analyzing the costs associated with misclassified cases. Future research could train models using all graduates with a sliding window method to account for year-over-year sequential changes.

4.2 | Utah System of Technical Colleges

Though USHE and UTech each serve different demographics of students with unique objectives, this study extracted many commonalities between the two groups. Namely, graduates who originate from Utah exhibit workforce retention rates higher than the cohort mean, previous employment history prior to graduation is a significant indicator of retention, and degree type/award level appears to influence an individual's economic opportunities. In total, eight out of ten UTech graduates were retained one year after certificate completion, and seven out of ten were retained five years after.

UTech's workforce retention rates differed from USHE, which was expected due to the differing demographics of an average UTech student compared to a USHE student. UTech had a higher percentage of students who were working prior to their enrollment compared to USHE, and UTech also had more students originating from Utah with 96% of the cohort being from in state compared to 71% of USHE students.

A primary trend in the UTech cohort data is the significantly lower longitudinal retention rates among those of minority race or ethnic groups. White graduates accounted for the vast majority of the UTech population, and their workforce retention rates exceeded the cohort average when analyzing both short- and long-term certificates.

Conversely, all graduates of minority race groups experienced workforce retention rates well below the cohort average, with many of these rates being statistically significant. This effect can be noticed when evaluating workforce retention at both the one- and five-year increments and raises questions about the inclusivity of Utah's economy. Workforce retention rates by race/ethnicity were informed by an interaction with geographical origin, with White and Hispanic in-state students having higher retention rates than their out-of-state counterparts (Appendix G).

Additionally, graduates from specific institutions (especially those who completed certificates at Davis Technical College and Uintah Basin Technical College) had lower workforce retention rates across time when compared to their counterparts. Uintah Basin Technical College's lower workforce retention rates are likely due to a majority of its graduates becoming self-employed as a consultant or through their Commercial Driver's License (CDL) program, which would not show up in the UI wage records. For Davis Technical College, the low workforce retention rate may be from those employed by Hill Air Force Base, who have high mobility across state lines for employment. It may also be because federal employees, which are considerable in Davis County, do not report wages to UI and are not captured in the data. In addition, graduates from Davis Technical College may go on to enroll in further higher education and do not show up in the wage records since they are continuing their education instead of entering the workforce.

Assessing the pattern between workforce retention and CIP code also suggested that certificates embracing STEM principles (e.g., health professions and related clinical sciences, engineering technology) may be more economically viable and contribute to higher retention rates. This localized relationship supports the nationwide trend of STEM career opportunities rapidly expanding, but also signifies a potential economic displacement of individuals who do not necessarily pursue technical/STEM education.

Holistically, other demographic variables like gender and age had minimal impact on workforce retention at both time increments. However, graduates who did not specify their gender had statistically significant lower retention rates. Although these individuals accounted for a small fraction of the UTech cohort (0.23%), further research could assess the connection between gender and economic opportunities.

Similar to USHE, the UTech analysis identified a relationship between workforce retention and the award level of the certificate attained (e.g., the number of requisite hours needed for completion).

Though not statistically significant, graduates with certificates requiring beyond 1,800 hours had steady workforce retention rates below the cohort mean. This may suggest additional education is correlated with a higher likelihood of departing Utah. However, technical or vocational certificates may not have the same level of portability as awards conferred from traditional universities, and additional information may be warranted to more comprehensively understand this effect. Another explanation may be this group enters self-employment after graduation and, therefore, do not appear in the UI wage records. Self-employment is likely for this group as 88% of UTech's long-term certificate graduates come from cosmetology programs.

Reaffirming the results from the USHE cohort, and in alignment with Koehler et. al (2017), this UTech analysis determined various permutations of individual variables can be leveraged to predict future workforce retention with reasonable accuracy. This analysis was also subject to losses in accuracy attributable to bias toward Type 1 errors. These errors inherently presume graduates will be retained and predict retention rates that are higher than reality. However, these predictive models did achieve reasonable accuracy and illuminate correlations between workforce retention and student characteristics.

4.3 | Limitations

One of the largest limitations of the study was the inability of wage records to capture some forms of employment. Self-employment, agricultural, federal, and non-profit employment are among the categories that may not be subject to UI requirements. While these results may not be generalized to omitted industries or employment types, the U.S. Bureau of Labor Statistics (2018) states that more than 90% of workers are included in the records, which still makes the report findings significant.

Additionally, the wage records provide no detail on other excluded individuals. Whether an employee retires, dies, or otherwise leaves the labor force or state cannot be determined using the data. However, given the age limit used in this study is restricted to students born after 1973, it is likely that few in the cohort would retire or die. Further research would be required to gain more insight into graduates who voluntarily leave the labor force or state.

Another limitation of the study occurred in CIP codes for graduates who received multiple awards at one time — a total of 408 individuals. Due to lack of feasibility in knowing which CIP code will directly lead to employment or which one is preferred by the student, a CIP code was selected

at random. A Pearson correlation was performed to test for significance of earning multiple USHE degrees on workforce retention. The results showed no significant correlation (r(22,071) = -0.022, $p <$ 0.001) between holding multiple awards and five-year retention. Given this, there is little statistical concern with the exclusion of multiple additional awards.

Data used for the study also provides an incomplete picture of all graduates in the state of Utah. UDRC does not have information on students who live in the state but take courses online at an out-of-state or non-USHE institution such as Brigham Young University, Westminster College, and Western Governors University. Therefore, data are only generalizable to graduates of USHE institutions.

As this study was carried out using the 2013 cohort, the prediction models may not have captured latent variables, which contributed to the workforce retention rate over time. For example, employment opportunities available and housing costs are factors many graduates take into consideration. As an example, those who graduated in 2013 may have faced different job and housing markets from those who graduated in 2016. As the scope of this study focused on the characteristics of the graduates, these hidden variables were not included as features. Further research with Long Short-Term Memory networks (LSTM) may capture changes in the external environment as well as characteristics of the graduates.

Lastly, the study was limited in its analysis of the interaction between CIP choice and geographic origin among UTech graduates. As previously mentioned regarding the USHE cohort, these specific results would not be publishable in accordance with UDRC disclosure protocol to preserve the identities of groups with fewer than 10 graduates.

5 | Conclusion

Ultimately, retention of recent graduates builds a well-educated workforce. By analyzing the characteristics of graduates who choose to remain in Utah, this study can help the state better understand who would most likely become future contributors to Utah's economy. To determine the workforce retention rates of graduates, in-state employment outcomes were analyzed for one year and five years after a USHE or UTech graduation in 2013. When comparing the one- and five-year retention rates, overall workforce retention rates were higher at one year after graduation and decreased at five years. Characteristics that influenced workforce retention rates in the analysis included geographic origin, gender, age, field of

study, award level, graduating institution, and financial aid.

Findings show that USHE and UTech graduates who originated from in-state and worked prior to enrollment were most likely to join the state's workforce after graduation. For those students who leave after graduation, out-of-state origin and higher education levels were correlated with a higher likelihood of departure.

Indicator variables such as geographic origin, field of study, and institution of graduation were used in gradient boosted decision trees and artificial neural networks to predict the retention rates of the 2016 graduating cohorts. When looking five years ahead, the models predicted that 70.76% of USHE's graduates and 72.63% of UTech's graduates would be retained in Utah's workforce in 2021.

While workforce retention of Utah's postsecondary graduates was the focus of this research, it is only one part of the overall equation in building a well-educated workforce. A full understanding of this equation would require additional research on the migration of educated workers, regardless of whether they received an education in Utah or out of state.

Data Partners

References

Fayer, S., Lacey, A., & Watson, A. (2017). STEM Occupations: Past, Present, And Future. Spotlight on Statistics, 7. https://www.bls.gov/spotlight/2017/science-technology-engineering-and-mathematics-stem-occupations-past-present-and-future/pdf/science-technology-engineering-and-mathematics-stem-occupations-past-present-and-future.pdf

Groen, J. A. (2004). The effect of college location on migration of college-educated labor. Journal of Econometrics, 121(1–2), 125–142. https://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1040&context=workingpapers

Harrington, J., Muñoz, J., Curs, B., & Ehlert, M. (2016). Examining the Impact of a Highly Targeted State Administered Merit Aid Program on Brain Drain: Evidence from a Regression Discontinuity Analysis of Missouri's Bright Flight Program. Research in Higher Education, 57(4), 423–447. https://doi-org.libdata.lib.ua.edu/10.1007/s11162-015-9392-9

Ishitani, T. (2011). The Determinants of Out-Migration Among In-State College Students in the United States. Research in Higher Education, 52(2), 107–122. https://doi-org.libdata.lib.ua.edu/10.1007/s11162-010-9187-y

Kodrzychi, Y. K. (2001). Migration of recent college graduates: Evidence from the national longitudinal survey of youth. New England Economic Review, January/February, 13–22. https://pdfs.semanticscholar.org/4497/a1d61b95cb3d48c0b6d007581c49dee7bb13.pdf

Koehler, T. J., Goodfellow, J., Davis, A. T., Spybrook, J., vanSchagen, J. E., & Schuh, L. (2017). Predicting In-State Workforce Retention After Graduate Medical Education Training. Journal of Graduate Medical Education, 9(1), 73–78. https://doi-org.libdata.lib.ua.edu/10.4300/JGME-D-16-00278.1

Leguizamon, J. S., & Hammond, G. W. (2015). Merit-based college tuition assistance and the conditional probability of in-state work. Papers in Regional Science, 94(1), 197–218. https://doi-org.libdata.lib.ua.edu/10.1111/pirs.12053

U.S. Bureau of Labor Statistics (2018). QCEW UI Overview. Retrieved from https://www.bls.gov/cew/cewover.htm

Yazback, S. (2005). Losing Its Minds? Evaluating "Brain Drain" in Ohio. Economic Commentary, 1–4. Retrieved from https://search-ebscohost-com.libdata.lib.ua.edu/login.aspx?direct=true&db=aph&AN=15954860&site=ehost-live

Appendix A

Hyperparameters for final machine learning models

USHE

Gradient Boosted Decision Tree		
	1-year	5-year
n_estimators	80	100
learning_rate	0.2	0.2
max_depth	3	3
Kernel SVM		
	1-year	5-year
kernel	'rbf'	'rbf'
C	1	2
gamma	0.1	0.1
k-Nearest Neighbors		
	1-year	5-year
n_neighbors	13	20
metric	'minkowski'	'minkowski'
p	2	2
Artifical Neural Network (MLP)		
	1-year	5-year
solver	'adam'	'adam'
alpha	0.05	0.05
max_iter	1000	1000
hidden_layer_sizes	(25, 25, 25)	(25, 25, 25)
learning_rate	'constant'	'constant'
activation	'tanh'	'tanh'

Gradient Boosted Decision Tree		
	1-year	5-year
n_estimators	100	100
learning_rate	0.001	0.001
max_depth	1	1
Kernel SVM		
	1-year	5-year
kernel	'rbf'	'rbf'
C	1	1
gamma	auto'	auto'
k-Nearest Neighbors		
	1-year	5-year
n_neighbors	5	5
metric	'minkowski'	'minkowski'
p	2	2
Artifical Neural Network (MLP)		
	1-year	5-year
solver	'adam'	'adam'
alpha	0.05	0.05
max_iter	100	100
hidden_layer_sizes	(5, 5, 5)	(5, 5, 5)
learning_rate	'constant'	'constant'
activation	softmax'	softmax'

USHE - One-year workforce retention rates by 2012 employment status and award level

	% Retained 1 Year
ASSOCIATE'S DEGREE	
Did Not Work in 2012	39.37%
Worked in 2012	82.52%
BACHELOR'S DEGREE	
Did Not Work in 2012	42.04%
Worked in 2012	82.33%
CERTIFICATE (<900h)	
Did Not Work in 2012	40.20%
Worked in 2012	91.05%
CERTIFICATE (>=900h, <1800h)	
Did Not Work in 2012	46.30%
Worked in 2012	89.84%
DOCTORATE - PROFESSIONAL PRACTICE	
Did Not Work in 2012	38.03%
Worked in 2012	75.73%
DOCTORATE - RESEARCH	
Did Not Work in 2012	25.42%
Worked in 2012	49.09%
MASTER'S DEGREE	
Did Not Work in 2012	36.97%
Worked in 2012	81.87%
POST-BACCALAUREATE CERTIFICATE	
Did Not Work in 2012	44.68%
Worked in 2012	80.77%
POST-MASTER'S CERTIFICATE	
Did Not Work in 2012	50.00%
Worked in 2012	81.08%

USHE - One-year workforce retention rates by 2012 employment and geographic origin

	% Retained 1 Year
IN-STATE	
Did Not Work in 2012	52.24%
Worked in 2012	84.70%
INTERNATIONAL	
Did Not Work in 2012	14.00%
Worked in 2012	56.42%
OUT-OF-STATE	
Did Not Work in 2012	28.93%
Worked in 2012	74.38%
UNKNOWN	
Did Not Work in 2012	42.05%
Worked in 2012	74.16%

USHE - One-year workforce retention rates by interaction of race/ethnicity and geographic origin

Race/Ethnicity	Origin	% Retained 1-Year
American Indian/Alaskan Native	In-state	67.31%
American Indian/Alaskan Native	Out-of-state	35.56%
American Indian/Alaskan Native	Unknown	N<10
Asian	In-state	76.59%
Asian	International	49.09%
Asian	Out-of-state	41.98%
Asian	Unknown	N<10
Black/African American	In-state	78.26%
Black/African American	International	83.33%
Black/African American	Out-of-state	35.78%
Black/African American	Unknown	N<10
Hispanic/Latino	In-state	75.49%
Hispanic/Latino	International	62.86%
Hispanic/Latino	Out-of-state	40.93%
Hispanic/Latino	Unknown	N<10
Multiple	In-state	79.61%
Multiple	International	N<10
Multiple	Out-of-state	56.79%
Multiple	Unknown	N<10
Native Hawaiian/Pacific Islander	In-state	80.25%
Native Hawaiian/Pacific Islander	Out-of-state	41.46%
Native Hawaiian/Pacific Islander	Unknown	N<10
Non Resident Alien	In-state	32.65%
Non Resident Alien	International	22.27%
Non Resident Alien	Out-of-state	N<10
Non Resident Alien	Unknown	N<10
Unspecified	In-state	72.58%
Unspecified	International	N<10
Unspecified	Out-of-state	62.66%
Unspecified	Unknown	N<10
White	In-state	77.77%
White	International	51.02%
White	Out-of-state	59.07%
White	Unknown	68.27%

USHE - Confusion matrices for machine learning models

Gradient Boosted Decision Tree		
	Predicted: Not-retained	Predicted: Retained
Actual: Non-retained 1-Year	794	910
Actual: Retained 1-Year	304	3,511
Actual: Non-retained 5-Year	1,409	868
Actual: Retained 5-Year	353	2,889
Kernel SVM		
	Predicted: Not-retained	Predicted: Retained
Actual: Non-retained 1-Year	675	1,029
Actual: Retained 1-Year	246	3,569
Actual: Non-retained 5-Year	1,375	902
Actual: Retained 5-Year	333	2,909
k-Nearest Neighbors		
	Predicted: Not-retained	Predicted: Retained
Actual: Non-retained 1-Year	847	857
Actual: Retained 1-Year	417	3,398
Actual: Non-retained 5-Year	1,435	842
Actual: Retained 5-Year	458	2,784
Artifical Neural Network (MLP)		
	Predicted: Not-retained	Predicted: Retained
Actual: Non-retained 1-Year	648	1,056
Actual: Retained 1-Year	195	3,620
Actual: Non-retained 5-Year	1,374	903
Actual: Retained 5-Year	321	2,921

Characteristics of the 2016 USHE cohort

	n	% of cohort
COHORT	29,351	100.00%
ORIGIN		
In-State	19,984	68.09%
Rural	2,919	9.95%
Urban	15,857	54.03%
Unknown county	1,208	4.12%
Out-of-State	7,517	25.61%
International	1,515	5.16%
Unknown	335	1.14%
EMPLOYMENT		
Worked in 2012	21,315	72.62%
Did not work in 2012	8,036	27.38%
DEMOGRAPHICS		
Sex		
Male	14,372	48.97%
Female	14,979	51.03%
Age		0.00%
Age: Under 29	23,548	80.23%
Age: 29 to 40	5,803	19.77%
Ethnicity		
American Indian/Alaskan Native	199	0.68%
Asian	757	2.58%
Black/African American	307	1.05%
Hispanic/Latino	2,047	6.97%
Multiple	589	2.01%
Native Hawaiian/Pacific Islander	136	0.46%
Non-Resident Alien	1,236	4.21%
Unspecified	1,317	4.49%
White	22,763	77.55%

Characteristics of the 2016 USHE cohort

	n	% of Cohort
AWARD LEVEL		
Associate's Degree	9,803	33.40%
Bachelor's Degree	14,330	48.82%
Certificate (<900h)	487	1.66%
Certificate (>=900h, <1800h)	798	2.72%
Doctorate - Professional Practice	311	1.06%
Doctorate - Research	243	0.83%
Master's Degree	3,017	10.28%
Post-Baccalaureate Certificate	263	0.90%
Post-Master's Certificate	99	0.34%
INSTITUTION		
Dixie State University	1,621	5.52%
Salt Lake Community College	4,159	14.17%
Snow College	848	2.89%
Southern Utah University	1,707	5.82%
University of Utah	6,870	23.41%
Utah State University	5,583	19.02%
Utah Valley University	4,354	14.83%
Weber State University	4,209	14.34%
FINANCE		
Loans		
Accepted loans (2012)	8,481	28.90%
Did not accept loans (2012)	20,870	71.10%
Financial Aid		
Received financial aid (2012)	14,708	50.11%
Did not receive financial aid (2012)	14,643	49.89%

Utech - One-year retention rates by 2012 employment status and award level

	Did Not Work in 2012	Worked in 2012
Certificate (<900h)	60.52%	88.98%
Certificate (>=900h, <1800h)	58.82%	90.93%
Certificate(>=1800h)	71.21%	77.78%

Utech - One-year retention rates by 2012 employment and geographic origin

	Did Not Work in 2012	Worked in 2012
In-State	62.87%	89.36%
Out-Of-State	34.96%	74.70%

UTech - One-year workforce retention rates by interaction of race/ethnicity and geographic origin

Race/Ethnicity	Origin	% Retained 1-Year
American Indian/Alaskan Native	In-state	74.51%
American Indian/Alaskan Native	Out-of-state	80.00%
Asian	In-state	73.61%
Asian	Out-of-state	N<10
Black/African American	In-state	63.29%
Black/African American	Out-of-state	N<10
Hispanic/Latino	In-state	78.32%
Hispanic/Latino	Out-of-state	54.55%
Native Hawaiian/Pacific Islander	In-state	81.82%
Native Hawaiian/Pacific Islander	Out-of-state	N<10
Unknown	In-state	78.59%
Unknown	Out-of-state	45.95%
White	In-state	82.43%
White	Out-of-state	52.14%

UTech - Confusion matrices for machine learning models

Gradient Boosted Decision Tree		
	Predicted: Not-retained	Predicted: Retained
Actual: Non-retained 1-Year R0	8	201
Actual: Retained 1-Year R1	8	917
Actual: Non-retained 5-Year R0	14	275
Actual: Retained 5-Year R1	11	834
Kernel SVM		
	Predicted: Not-retained	Predicted: Retained
Actual: Non-retained 1-Year	0	209
Actual: Retained 1-Year	0	925
Actual: Non-retained 5-Year	0	289
Actual: Retained 5-Year	0	845
k-Nearest Neighbors		
	Predicted: Not-retained	Predicted: Retained
Actual: Non-retained 1-Year	8	201
Actual: Retained 1-Year	15	910
Actual: Non-retained 5-Year	25	264
Actual: Retained 5-Year	22	823
Artifical Neural Network (MLP)		
	Predicted: Not-retained	Predicted: Retained
Actual: Non-retained 1-Year	8	15
Actual: Retained 1-Year	201	910
Actual: Non-retained 5-Year	32	257
Actual: Retained 5-Year	20	825

APPENDIX I

Characteristics of the 2016 UTech cohort

	n	% of cohort
COHORT	5,933	100.00%
ORIGIN		
In-State	5,642	95.10%
Rural	1,113	18.76%
Urban	4,529	76.34%
Out-of-State	275	4.64%
Unknown	16	0.27%
EMPLOYMENT		
Worked in 2015	4,353	73.37%
Did not work in 2015	1,580	26.63%
DEMOGRAPHICS		
Sex		
Male	2,673	45.05%
Female	3,256	54.88%
Unspecified	1	0.02%
Age		
Age: Under 29	4,794	80.80%
Age: 29 to 40	1,009	17.01%
Unknown	130	2.19%
ETHNICITY		
American Indian/Alaskan Native	99	1.67%
Asian	108	1.82%
Black/African American	112	1.89%
Hispanic/Latino	749	12.62%
Native Hawaiian/Pacific Islander	51	0.86%
Unspecified	35	0.59%
White	4,847	81.70%
AWARD LEVEL		
Certificate (<900h)	3,497	58.94%
Certificate (>=900h, <1800h)	1,393	23.48%
Certificate (>=1800)	7	0.12%
Unknown Award	1,036	17.46%

Characteristics of the 2016 UTech cohort

	n	% of Cohort
INSTITUTION		
Bridgerland Technical College	827	13.94%
Davis Technical College	1,415	23.85%
Dixie Technical College	329	5.55%
Mountainland Technical College	1,666	28.08%
Ogden-Weber Technical College	959	16.16%
Southwest Technical College	307	5.17%
Tooele Technical College	147	2.48%
Uintah Basin Technical College	283	4.77%

F1-Scores and Matthews correlation coefficient of USHE and UTech Models

Model Name	Institution	Horizon	F1-Score	MCC
Gradient Boosted Decision Tree	UTECH	1	0.90	0.10
Kernel SVM	UTECH	1	0.90	NR
k-Nearest Neighbors	UTECH	1	0.89	0.06
Multi Layer Perceptron	UTECH	1	0.89	0.06
Gradient Boosted Decision Tree	UTECH	5	0.85	0.11
Kernel SVM	UTECH	5	0.85	NR
k-Nearest Neighbors	UTECH	5	0.85	0.13
Multi Layer Perceptron	UTECH	5	0.86	0.18
Gradient Boosted Decision Tree	USHE	1	0.85	0.45
Kernel SVM	USHE	1	0.85	0.41
k-Nearest Neighbors	USHE	1	0.84	0.43
Multi Layer Perceptron	USHE	1	0.85	0.42
Gradient Boosted Decision Tree	USHE	5	0.83	0.54
Kernel SVM	USHE	5	0.82	0.53
k-Nearest Neighbors	USHE	5	0.81	0.51
Multi Layer Perceptron	USHE	5	0.83	0.54