

UNIVERSITY *of* WASHINGTON

Dropout Prediction Project

Utilizing Data to Support Student Success



DANA
ABDIRAKHYM

DIANE
HOANG

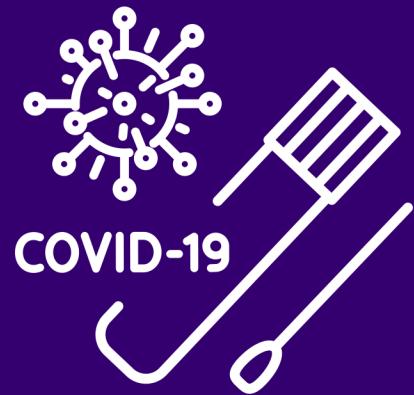
DOUNIA
BENJYDA

JUSTIN
CABANOS

SUKHDEEP
KAUR

WT

UNIVERSITY *of* WASHINGTON ORGANIZATIONAL CHALLENGES



“After 29 years of continuous year-over-year growth, Autumn 2019 marked the first decline in student enrollment, a trend that has persisted with the onset of COVID-19.”

“The University of Washington Tacoma has been experiencing an increasing number of undergraduate students leaving before completing their degrees and facing challenges in attracting new students.”

“Understanding the needs of students is fundamental to the continued growth and adaptability of UWT and our community. Currently, students are making educational decisions at a pace that outstrips UWT's ability to anticipate and respond effectively.”



UNIVERSITY *of* WASHINGTON SOLUTION ARCHITECTURE

TO ESTABLISH A APPLICATION THAT CAN HELP IDENTIFY WHEN A STUDENT IS AT RISK OF LEAVING, GIVING THE SCHOOL A CHANCE TO INTERVENE AND PROVIDE RESOURCES.

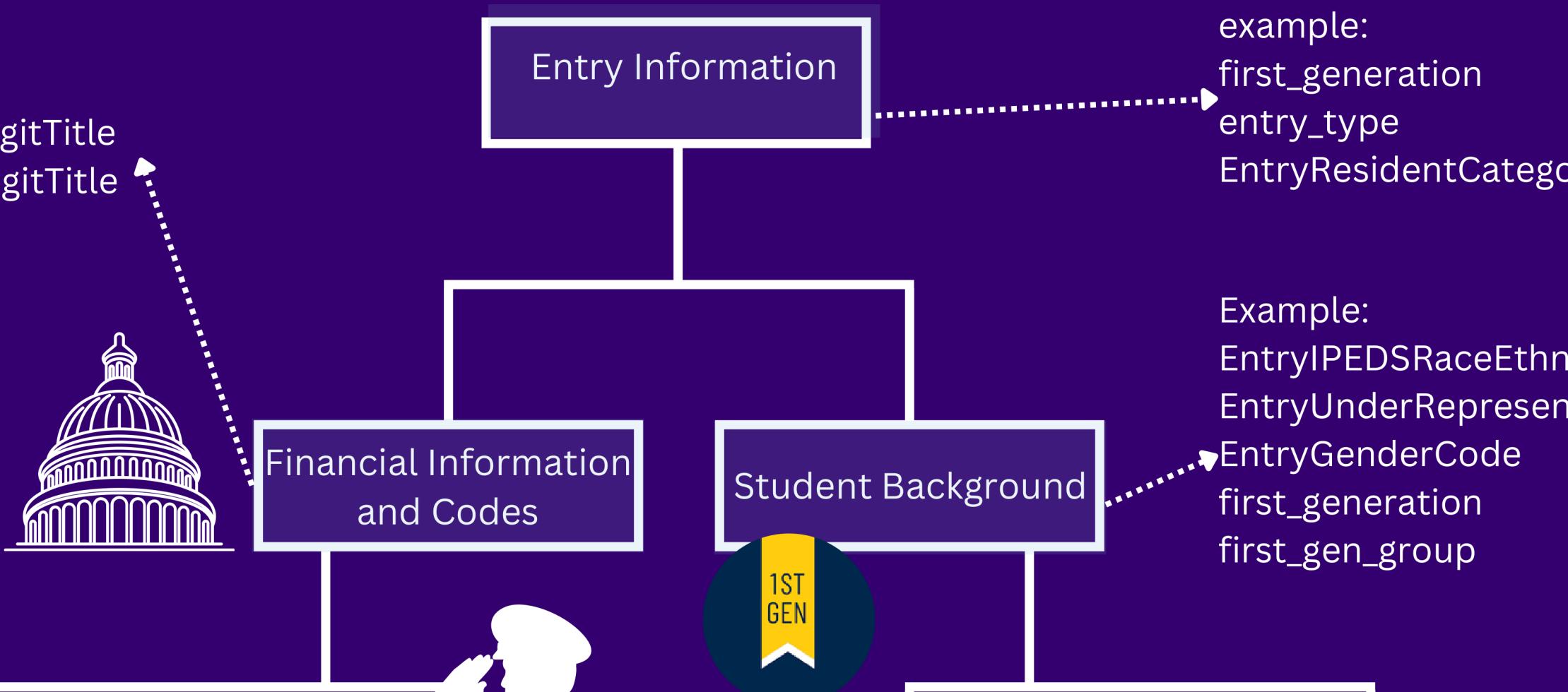
Increase Student Retention



UNIVERSITY of WASHINGTON

DATA OVERVIEW

Example:
CIPCode
CIPFamily2DigitTitle
CIPFamily4DigitTitle



example:
first_generation
entry_type
EntryResidentCategory

Source and
Identification Keys

Example:
EntryIPEDSRaceEthnicityCategory
EntryUnderRepresentedMinority
EntryGenderCode
first_generation
first_gen_group



Academic Performance
Demographics

Example:
CensusDayCumGPA

Veteran and Benefits
Information

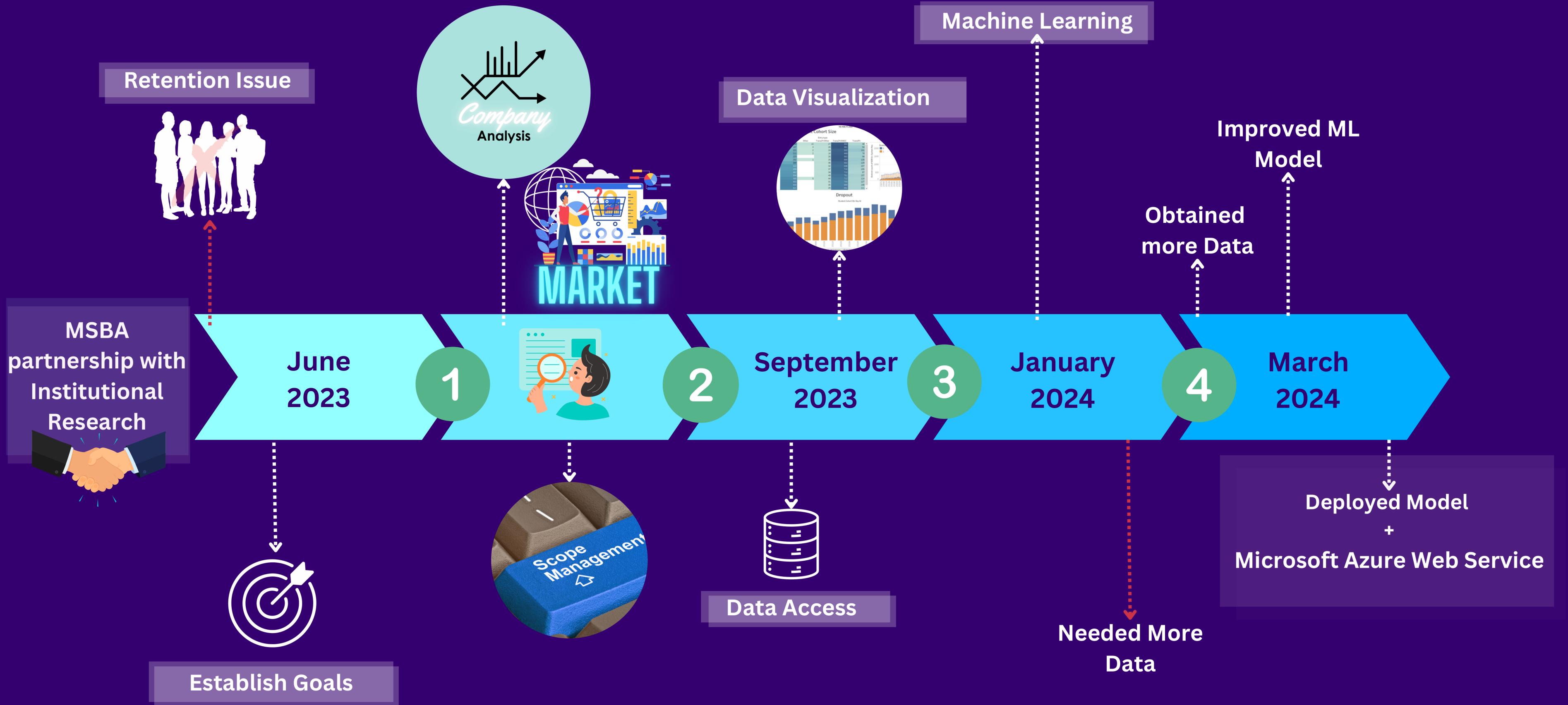
Example:
vet_benefit
veteran
veteran_descrip

Retention and
Graduation

Example:
degree_UWT
Graduated
TimeToDegreeInYears

Academic Progress and
Major Information

Example:
StudentCohortQtr
RetainedTerm01
Re_enrolled_next_academic_year



LIVE DEMO: DESCRIPTIVE ANALYTICS

WHICH STUDENTS HAVE HIGHER DROPOUT?

LOW GPA:

Below 1.2 = 100% dropout rate.

ACADEMIC YEAR:

Quarter 4 of 2022 had the highest dropout.

GENDER:

Males drop out more than females.

FIRST YEAR MAJORS:

Higher dropouts in 2 T HIST 70 and 2 TECON 10.

SECOND YEARS:

Higher dropouts in certain majors and colleges.

MILITARY DEPENDENTS:

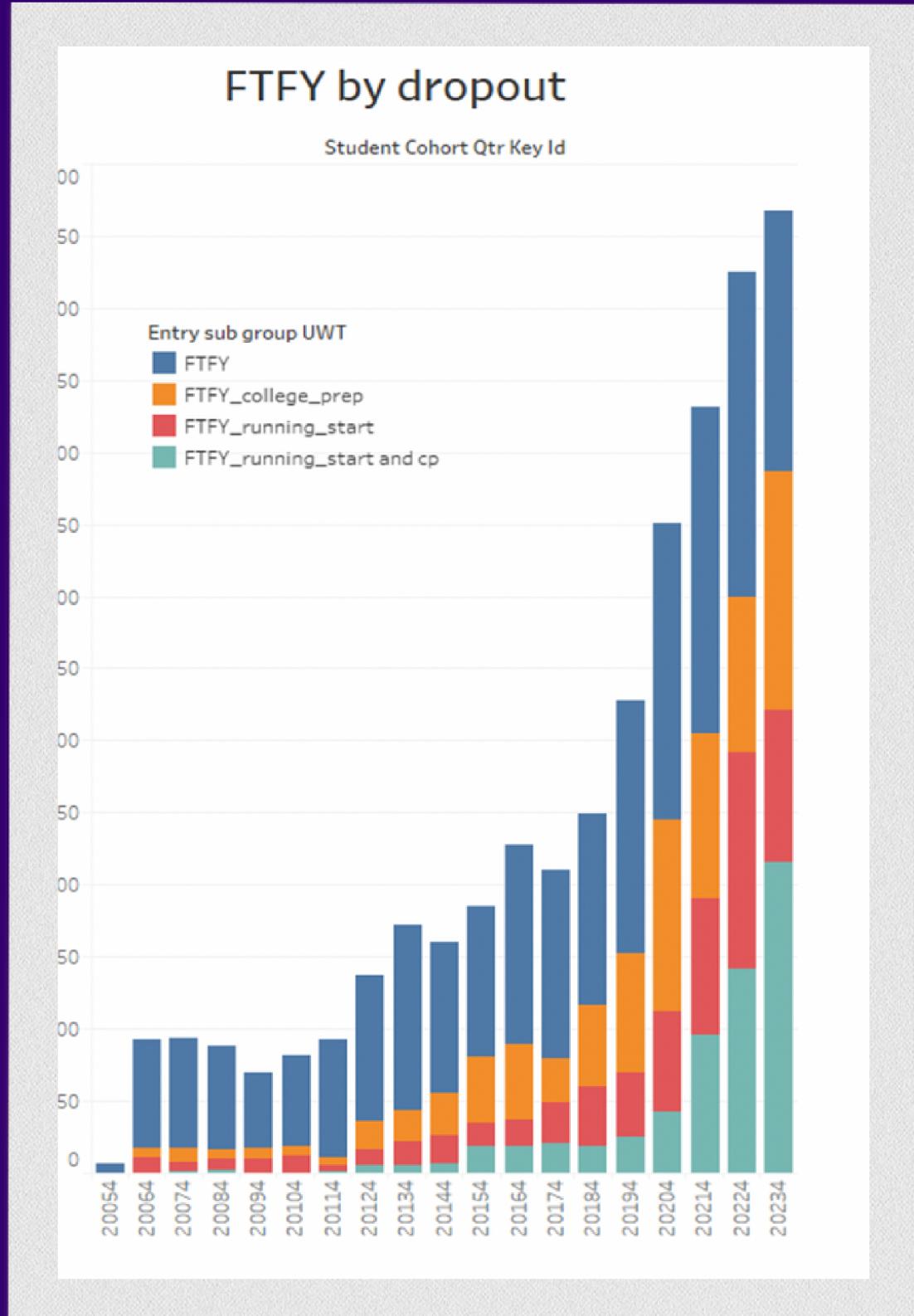
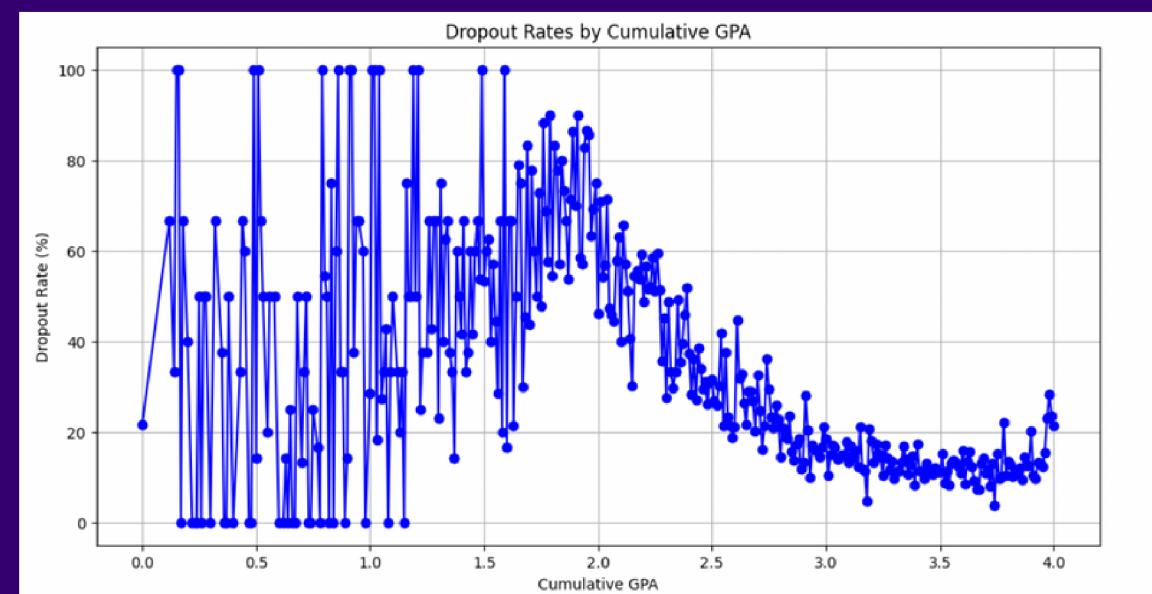
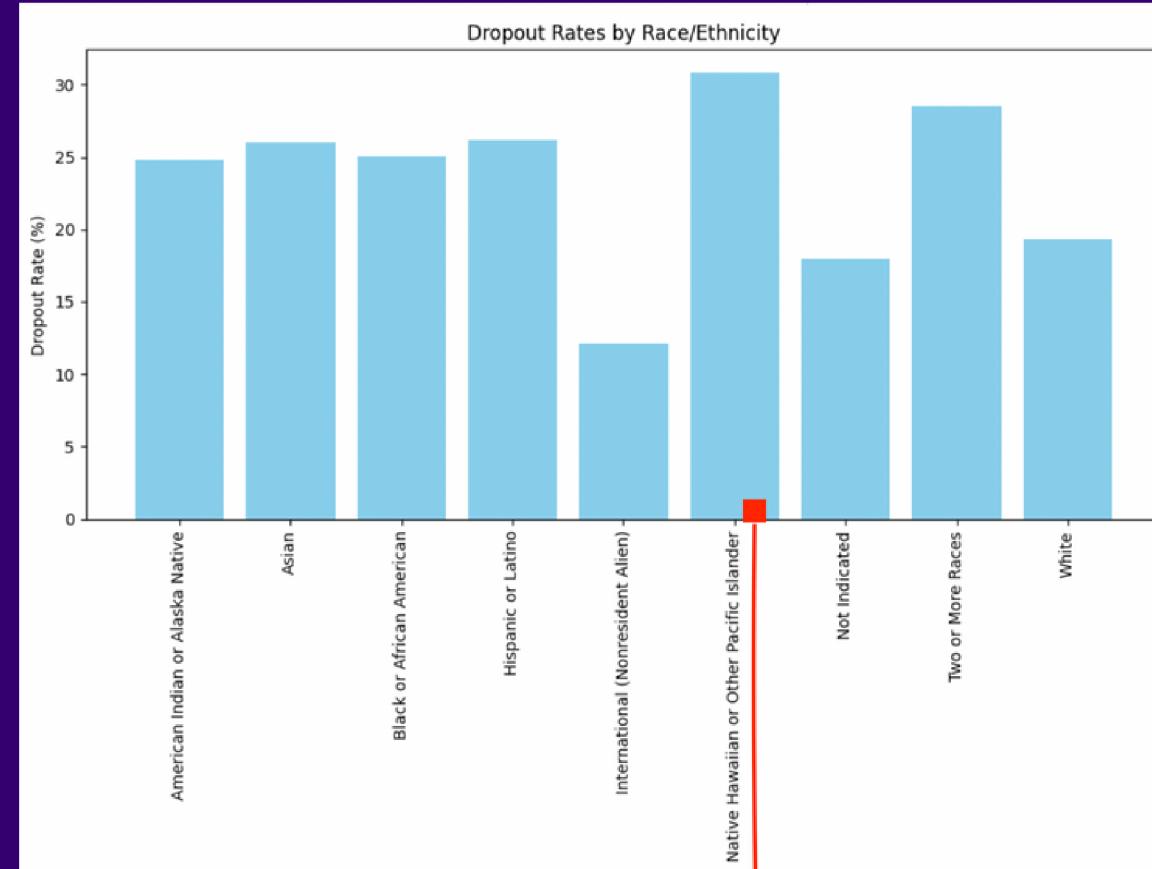
Higher dropout rate.

RACIAL/ETHNIC GROUPS:

Higher dropouts among Native Hawaiian, Pacific Islander, and Asian students.

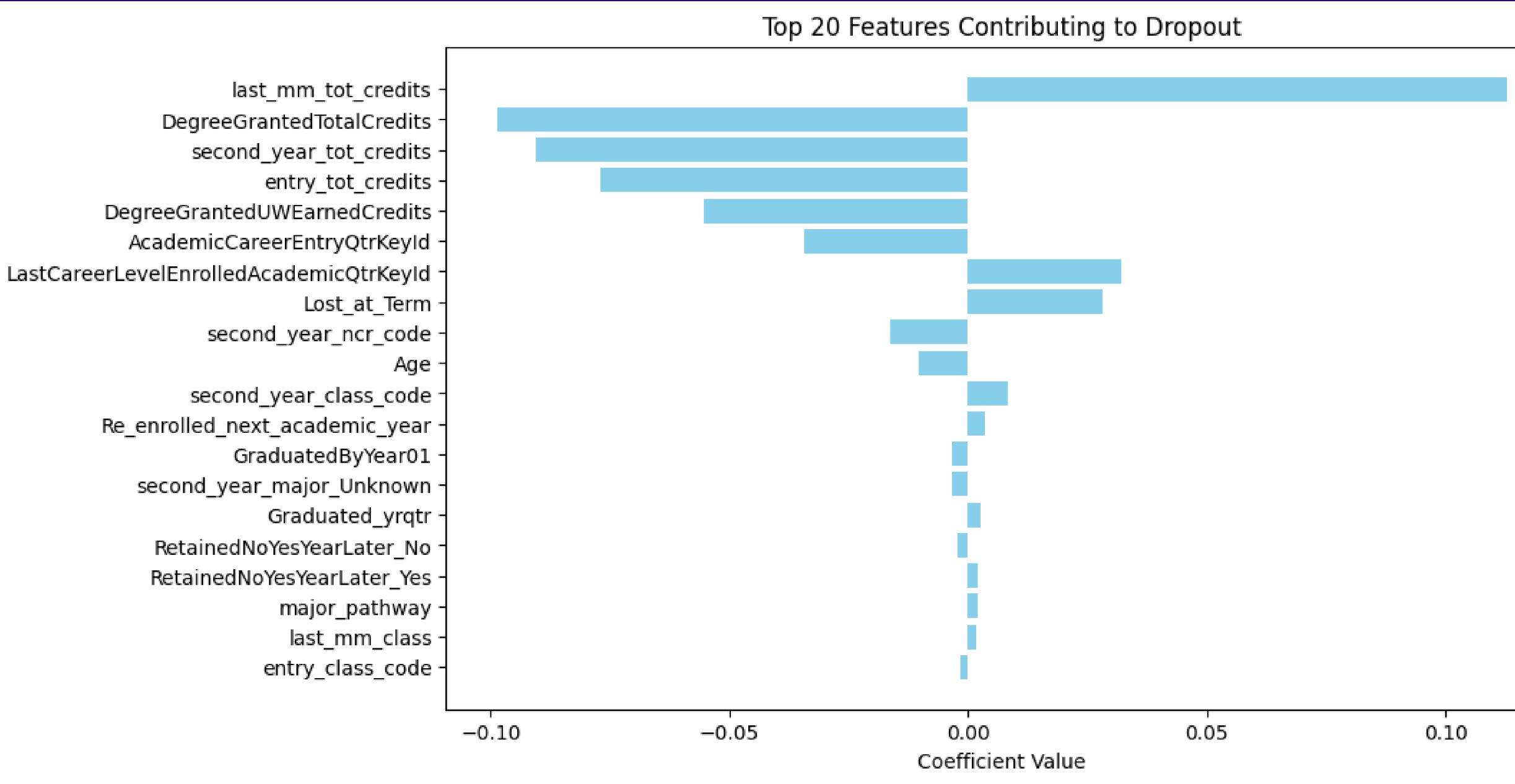
AGE:

Students aged 20-23 have a higher dropout rate.



LIVE DEMO: PREDICTIVE ANALYTICS

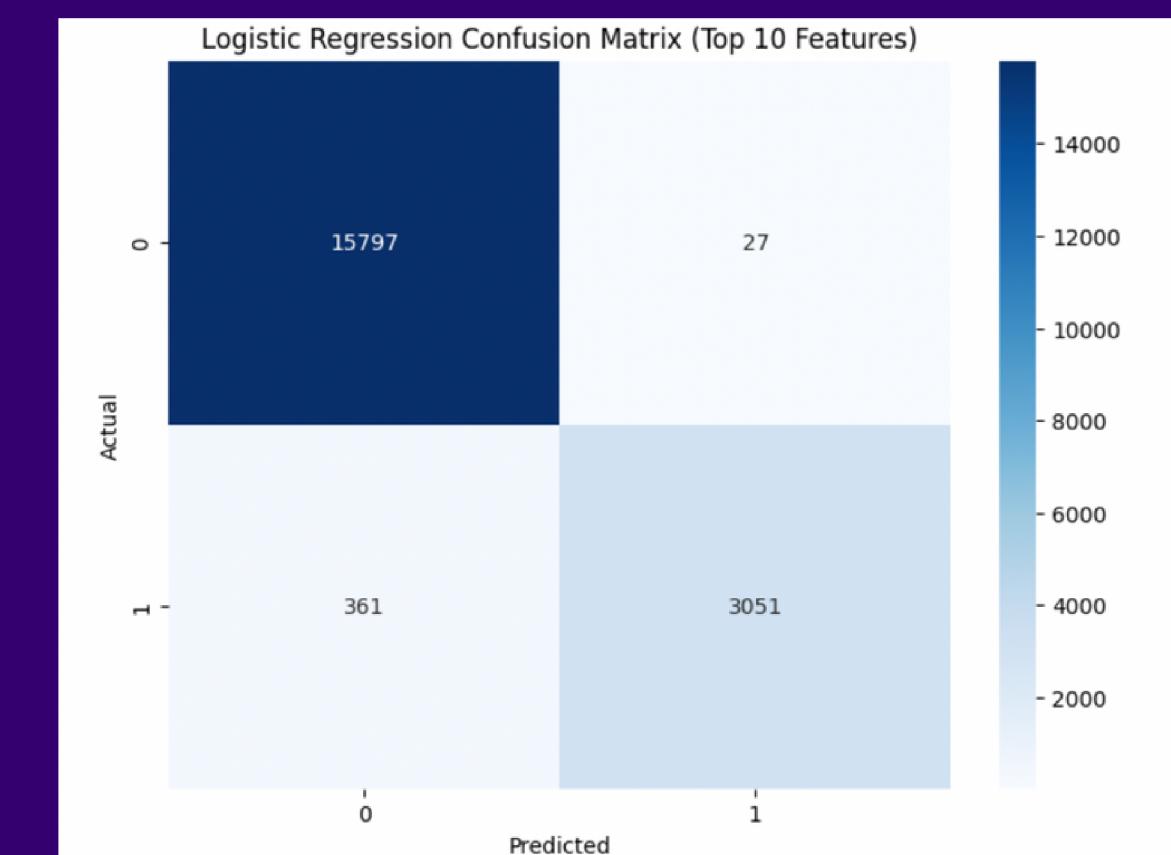
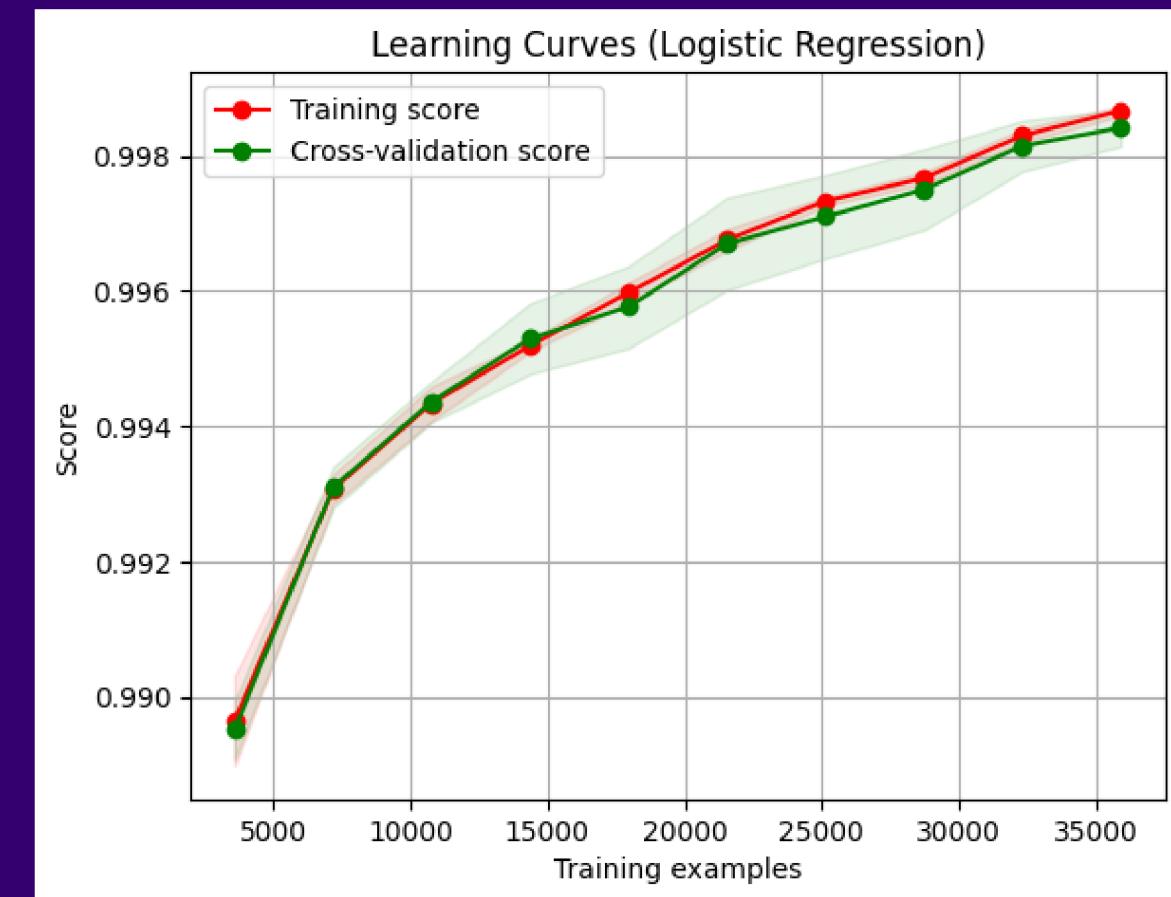
FEATURES THAT CONTRIBUTE TO DROPOUT



False Positives (FP): 27 - Students predicted to drop out but who do not.

False Negatives (FN): 361 - Students not predicted to drop out but who do, indicating a potential area for model improvement to capture more of these cases.

Cross-Validation Score: The learning curves converge and the performance gap between the training and validation sets is minimal. The small difference between the training and validation scores suggests that the model is not overfitting. Instead, it learns the data patterns effectively without becoming too tailored to the training data specifics.



LIVE DEMO: PREDICTIVE ANALYTICS

Last Term Total Credits (11.8141):

Students with more credits in their last term are significantly more likely to drop out.

Total Degree Credits (-7.80713):

Accumulating more credits towards a degree correlates with a lower dropout likelihood.

UW Earned Credits (-4.94404):

Earning more credits from the University of Washington reduces dropout probability.

UW Tacoma Degree (-4.58006):

Graduating from UW Tacoma is associated with a decreased dropout risk.

Transfer Student (-2.3259):

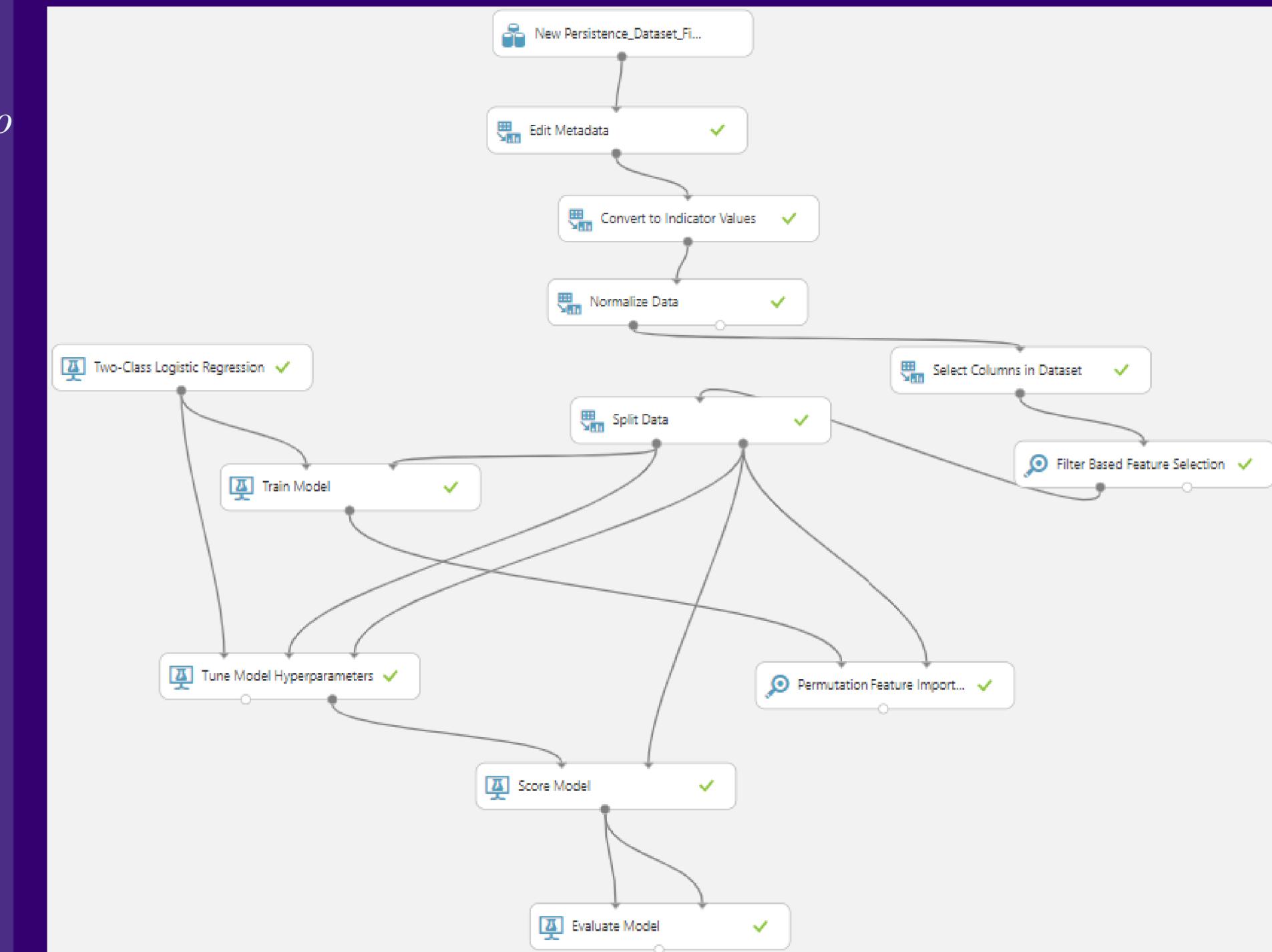
Transfer students are less likely to drop out compared to non-transfer students.

Time to Degree in Years (-2.11145):

Taking longer to complete a degree is linked with reduced dropout rates.

Age (-0.631151):

Older students have a slightly lower chance of dropping out.



LIVE DEMO: PRESCRIPTIVE ANALYTICS

WHAT SOLUTIONS CAN WE OFFER TO UWTIR?

Power App that requires user inputs and generates the scored probability of dropout from Azure ML Model.
The app generates 0 if the student will graduate (green), 1 if the student drops out (red)

The image shows a Power Automate flow on the left and a Power App interface on the right.

Power Automate Flow:

```
graph TD; A[PowerApps (V2)] -- "0s ✓" --> B[HTTP]; B -- "0.2s ✓" --> C[Parse JSON]; C -- "0s ✓" --> D[Respond to a Power App or flow]
```

Power App Interface - Welcome to Student Dropout Prediction App:

Transfer student: 0
Degree UWT: 1
Total Credits: 100
UW Earned Credits: 70
Time to Degree in Years: 4
Last mm total credits: 0
Last mm class: 3
Age: 30

Predict

Scored Label: **0** Probability: **0.00**

A horizontal slider at the bottom ranges from 0 to 1, with a green dot at 0 indicating the current score.

LIVE DEMO: PRESCRIPTIVE ANALYTICS

- The purpose of the app is to predict the student dropout probability based on Logistic Regression, built in Azure ML.
- it is interactive tool generates the result based on features (transfer student, degree received as University of Washington, number of total credits, number of total credits earned at UW, time to finish the degree, last minimaster credits, last minimaster class and age)



Welcome to Student Dropout Prediction App

Transfer student	0	Time to Degree in Years	4
Degree UWT	0	Last mm total credits	180
Total Credits	0	Last mm class	4
UW Earned Credits	0	Age	30

Predict

Scored Label: **1** Probability: **0.99**

A screenshot of the "Student Dropout Prediction App". The interface has a light blue header bar with the title. Below it is a form with four columns of input fields. The first column contains "Transfer student" (0), "Degree UWT" (0), "Total Credits" (0), and "UW Earned Credits" (0). The second column contains "Time to Degree in Years" (4), "Last mm total credits" (180), "Last mm class" (4), and "Age" (30). At the bottom left is a purple "Predict" button. To the right of the "Predict" button is a horizontal scale from 0 to 1, with a red dot at 0.99. Above the scale, the text "Scored Label: 1" and "Probability: 0.99" is displayed.



UNIVERSITY *of* WASHINGTON

CHALLENGES • ACTIONS

CLASS IMBALANCE:

To address the imbalance in our dataset for predicting dropout, we focused on the Graduated column instead of using lost_at_term. We had 20,000 records for Graduated = 0 and 40,000 for Graduated = 1. By randomly sampling the Graduated = 1 group down to 20,000 records, we created a balanced dataset with 20,000 rows for each category, ensuring effective model training.



CATEGORICAL HEAVY DATASET:

Our initial dataset contained 70 out of 82 columns were categorical.

Additional numeric variables such as Age and GPA columns were added to dataset to enhance the model's accuracy and utility.

NULL VALUES:

To address the challenge of handling numerical and categorical null values, we consulted with the client to understand the reasons behind these null entries. Based on their input, we either renamed or assigned appropriate values to the nulls. This approach preserved as much data as possible for analysis, rather than simply removing the null values during data cleaning.



UNIVERSITY of WASHINGTON

RECOMMENDATIONS • INSIGHTS



Power app can provide UWTIR to detect the probability of dropout and scored label based on top 8 features. It saves time and easy to use. Great tool to use for professors, faculty members, advisors and registrar office. **Tableau dashboard done in Q2 can be used to dynamically check the dropout trends(link in remote desktop).**

Recomendations to UWTIR

- **Balanced Course Loads:** Encourage students to maintain a balanced course load and GPA especially in their final term, to reduce stress and enhance their chances of graduation.
Suggestion would be students can take summer classes.
- **Enhance Institutional Support:** The at risk majors must be provided more support from TLC.
Suggestion would be to get more resources in TLC.
- **Support students with less academic experience:** Required check-ins with advisors, introduce to the resources and study groups.
Suggestion would be to get mentors

Next Steps / insights:

- Expanding the dataset with additional demographic and external data (personal and work life).
- In addition to metadata website, create wholesome data dictionary that includes all the variables in dataset.
- What student population (sophomore, FTFY, transfer) and when (quarter) the students are struggling with higher course load?
- Incorporate scheduling and course time information to the data analysis next year.
- Make sure the dataset is balanced 50% dropout students and 50% graduated students before model training.

UNIVERSITY *of* WASHINGTON

TEAM 7



DANA
ABDIRAKHYM



DIANE
HOANG



DOUNIA
BENJYDA



JUSTIN
CABANOS



SUKHDEEP
KAUR



Thank you

TWENTY