THE UNIVERSITY OF WASHINGTON:

STUDENT SUCCESS AT UWT

**TACOMA INSTITUTIONAL RESEARCH, Consulting Team 7:**

Dounia Benjdya, Sukhdeep Kaur, Dana Abdirakhym, Justin Cabanos

**Client:** Joe Lawless, Chief Strategy Officer

Alice Few, Institutional Analyst

**February 5th, 2024**

**Table of Contents**

1. Executive summary 3

2. Data Analytics Plan and Techniques 4

2.1 Analytics Paradigm 4

2.2 Descriptive and Diagnostics Analytics 4

2.2.1 The questions to be answered 5

2.2.2 Tools and Techniques 5

2.2.3 Findings 5

2.3 Predictive Analytics 7

2.3.2 Tools and Techniques 7

2.3.3 Plan ahead 9

3. A Framework for Agile Analytics 10

3.1 Perform Business Discovery 10

3.2 Perform Data Discovery 11

3.3 Prepare the Data 11

3.4 Model Data 11

3.5 Score And Deploy 13

3.6 Evaluate and Improve 14

4. Conclusions 15

**Executive summary**

1. The objective of our document is to provide a clear management plan for the data collected from the University of Washington Tacoma’s institutional research center in order to aid their goal in detecting early warning signals of a student’s potential drop out. All our data will be stored and secured using UWT’s and Federal Education Rights and Privacy Act’s guidelines.

Roles and responsibilities with the document have been equally distributed among our group members, ensuring that each task and role is given based not only on the member’s strengths but with the possibility of creating and learning new skill sets. Roles include a project manager to ensure organization and deadlines within the project are being met, data steward to ensure handling and management of data, a visualizer to handle charts, plans and data visualization and a data profiler, who will provide a deeper understanding of the data beyond the numerical sets.

Our document will be set up in multiple sets:

1. **Data Analytics Plan and Techniques**

● Analytics Paradigm

● Descriptive and Diagnostics Analytic

● Predictive Analytics

2. **A Framework for Agile Analytics**

● Perform Business Discovery

● Perform Data Discovery

● Prepare the Data

● Model Data

● Score And Deploy

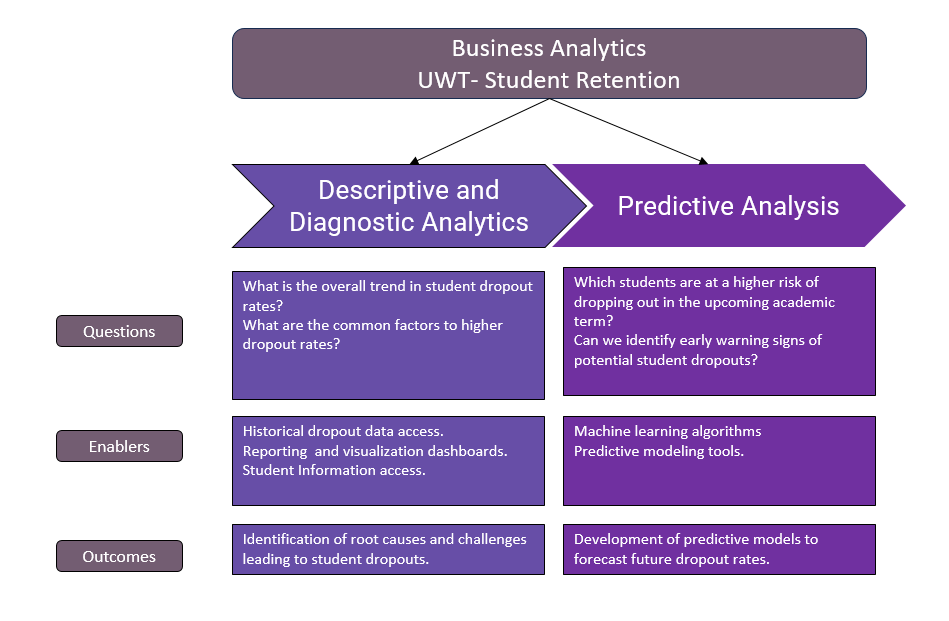
● Evaluate and Improve

3. **Conclusion**

**Data Analytics Plan and Techniques**

**2.1 Analytics Paradigm**

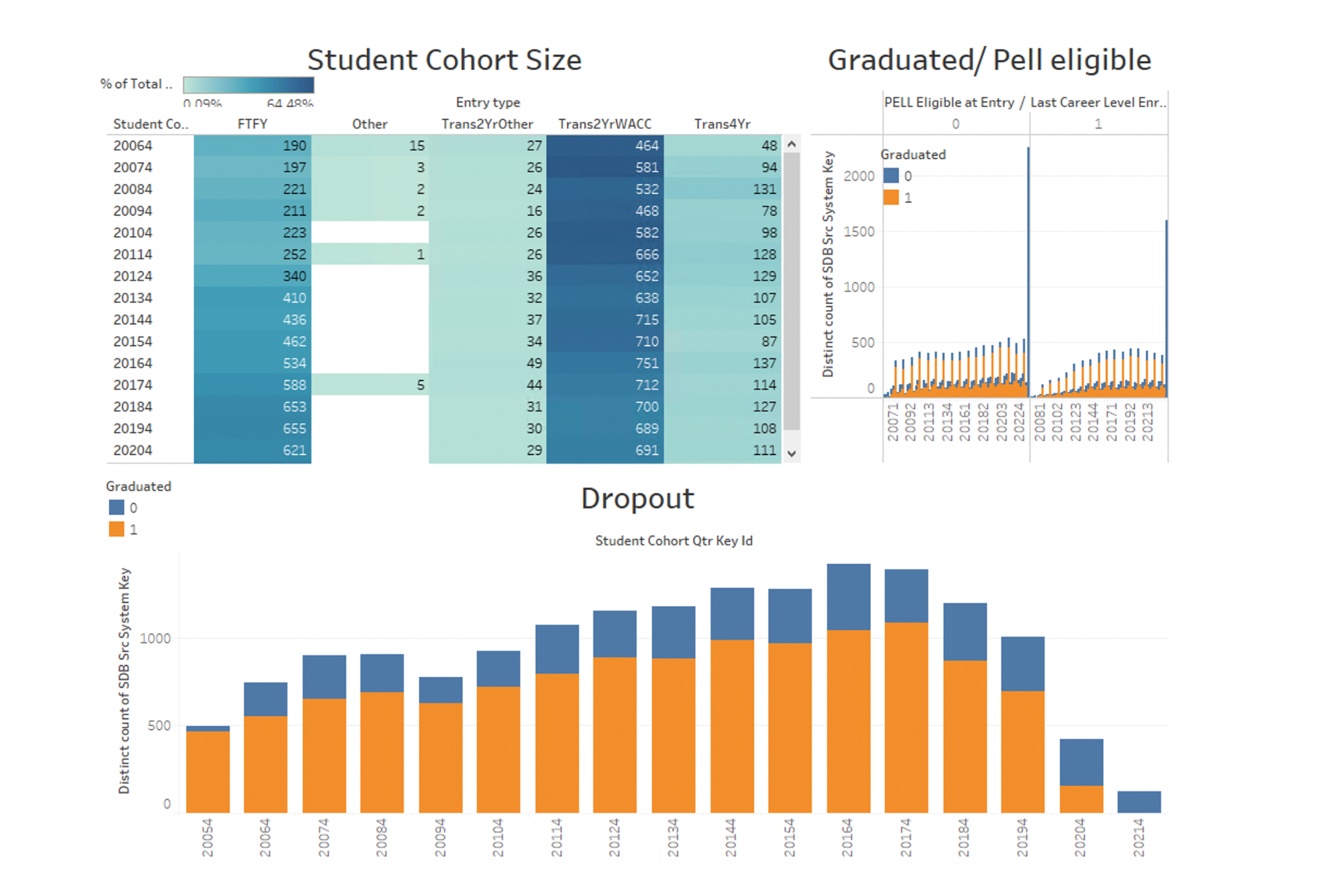
The analytical paradigm involves employing various analytical approaches and methodologies to understand the factors contributing to student dropout rates and to develop insights that can inform strategies for improvement. As a team we are working towards answering what factors are impacting student dropout rates and how to proactively identify at-risk students so the University of Washington, Tacoma (UWT) can turn a possible stop-out into a continued enrollment.



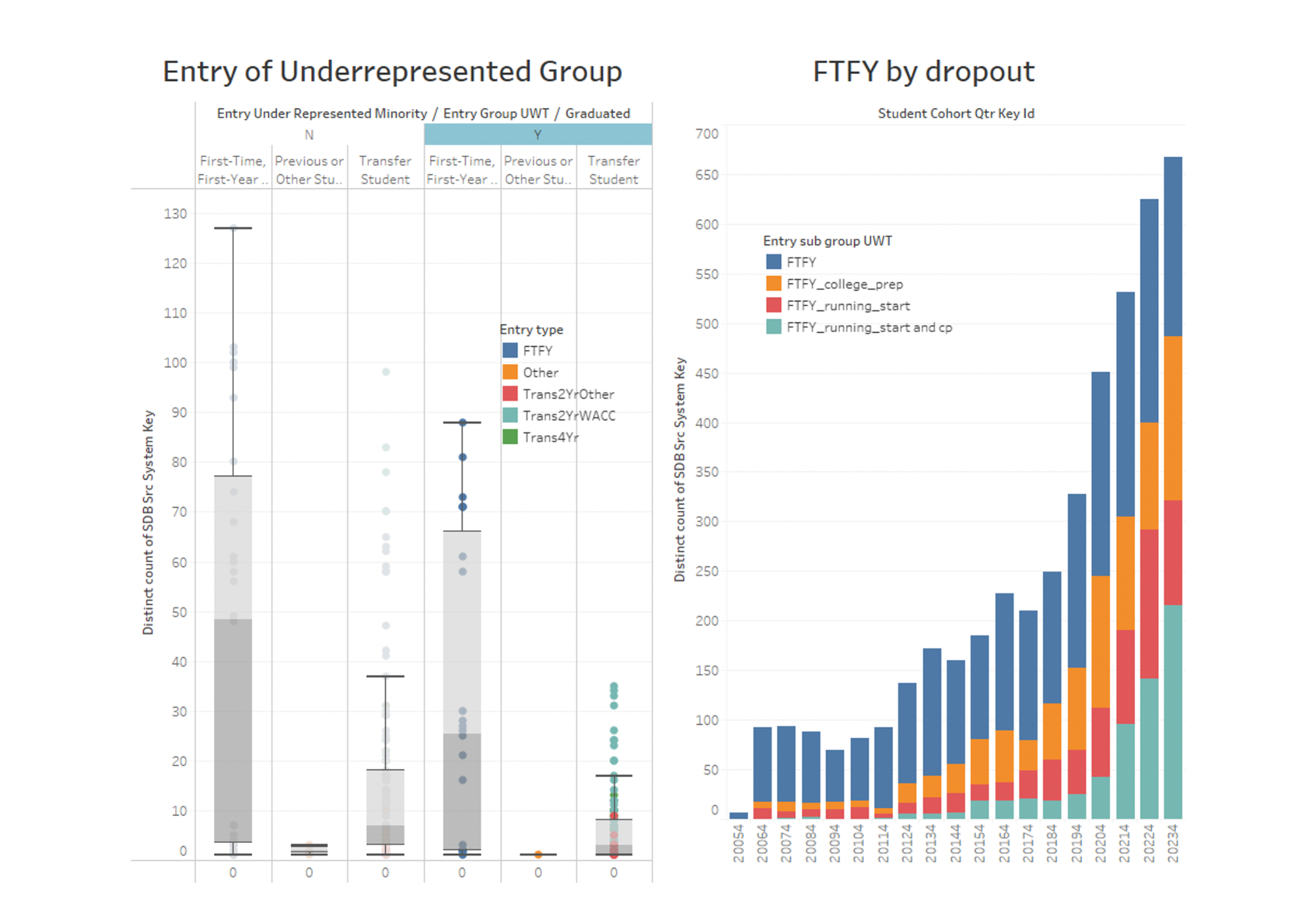
**2 Descriptive and Diagnostics Analytics**

Descriptive analytics allowed our team to summarize historical dropout data and identify trends and patterns over time in terms of how many students drop out of university. Diagnostic analytics involves investigation into why students have stopped out and acquired an understanding of the common challenges and problems experienced by dropouts, the reasons for the challenges, and the characteristics of stop-outs in terms of academic performance, attendance, and other socio-economic factors. By combining both descriptive and diagnostic paradigms, we can now explore ways to proactively identify students at risk, thereby turning a potential to be a stop-out into continuing enrolment and completion of a degree through appropriate intervention and support.

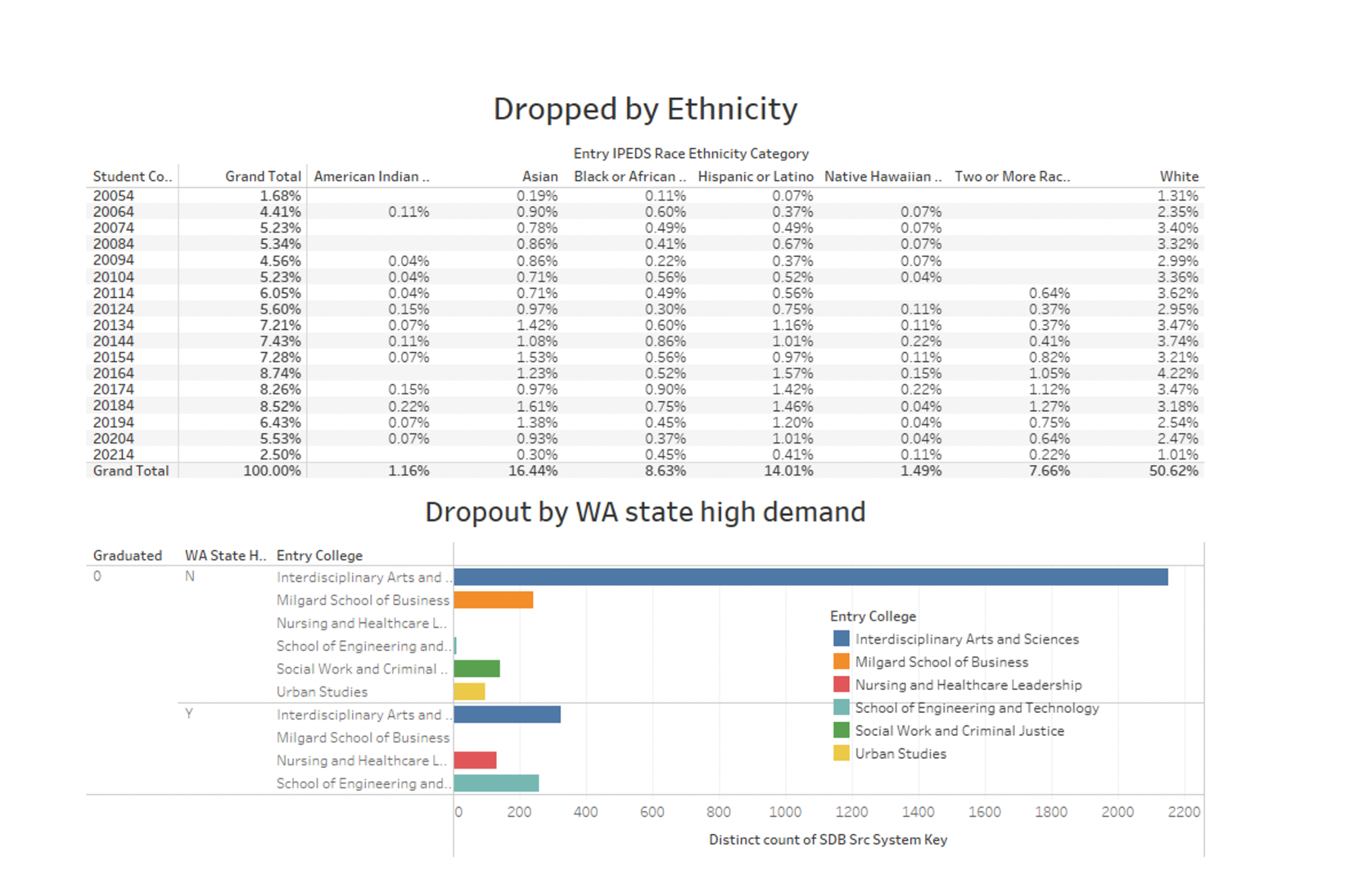
The diagnostic and descriptive approach to the variables involved was meticulous and systematic to uncover key insights into the intricate dynamics of the student dropout rate. First, the analysis focused on the relationship between the Pell grant eligibility and the dropout rate, using visualization techniques to identify patterns and relationships within the data. A granular analysis was then conducted to reveal a clear difference in the dropout rates between the eligible and ineligible cohorts. To understand the evolution of the dropout rates, an analysis was conducted from 2011 to 2020, revealing significant shifts in the undergraduate student attrition rates, especially during the tumultuous year of 2020 post COVID-19. Additionally, a detailed analysis of various student cohorts, including FTFY and highly-demand majors, revealed specific trends and challenges facing diverse student populations.



**Figure 1. Student attrition by cohort size and Pell grant eligibility.**



**Figure 2. First Time First Year students, Underrepresented group dropout rates.**



**Figure 3. Dropout rate by students’ ethnicity and Washington state high demand majors.**

**2.2.1 The questions to be answered**

1. What are the primary reasons behind student dropouts at UWT, and how do academic performance, attendance, and socio-economic factors contribute to these dropout rates?
2. How can at-risk students be proactively identified by UWT?
3. What sorts of targeted interventions and support mechanisms can be developed to keep potential stop-outs to continue to enroll at UWT?

**2.2.2 Tools and Techniques**

PowerBi, Microsoft Machine Learning, MS SQL Server, MS Excel, Tableau

**2.2.3 Findings**

Our overall findings discovered a steady increase of dropout rates by 5% throughout the past five years. The biggest increase happened in 2020, at the start of the pandemic. As provided from the tables shown in section 2.0, we conducted an analysis focusing on the count of student income levels, ethnicity, major and experience with higher education.

***Our findings were as followed:***

*Income:*

• Dropout and Ineligible for Pell: 5,864 students (10% of cohort)

• Dropout but Pell Eligible: 4,214 students (7.2% of cohort)

• Graduated, Not Pell Eligible: 10,968 students (18.7% of cohort)

• Graduated and Pell Eligible: 7,555 students (12.9% of cohort)

*Conclusion:*

Student income has no determination on a student’s determination to complete their degree at UWT

*Is their major in demand?* (Potential motivation to complete degree)

No: Interdisciplinary Art (2150 students)

Yes: Interdisciplinary Art (325 students)

No: Milgrad School of Business (242 students)

Yes: School of Engineering and Technology (257 students)

No: Social Work and Criminal Justice (142 students)

Yes: Nursing and Healthcare Leadership (131 students)

*Conclusion:*

Comparing the answers to the yes or no with the analysis we conducted above of dropout rates by major, we can see there is a connection between those that answered no and a higher dropout rate.

*Experience with higher education:*

• More than 50% of FTFY chose to drop out from AY 2016-2023

• Around 20% of FTFY\_college\_prep students started to dropout from 2020

• 15% FTFY\_running\_start student stopped attending UWT from 2022

• Almost 25% of Running start and college prep students decided to leave the UW Tacoma

*Conclusion:*

Students with no prior experience in higher education drop out at significantly higher rates than those who have had some variation of exposure to college level coursework.

*Race:*

• White: 42% drop in 2020 compared to 2016

• 2 or more Races: 30% drop

• Native Hawaiian: 1 student dropout compared to 3 in 2016

• Hispanic or Latina: 24% drop in dropout

• Black or African: 14 to 12 students

• Asian: 41 to 12

• American Indian: 2 students both years

*Conclusion:*

There seems to be a direct relation in the count of race of students who dropped out of the university in comparison to their peers, however, at this time it is hard to determine the reason why until we compare ratios.

**2.3 Predictive Analytics**

The next goal in our analysis would be to examine what is the statistical significance of whether the student drops out. Using predictive analytics, we will answer the following questions revolving around what will happen and how it will happen. We will use the Azure ML to implement our analysis and use the logistic regression to measure the output target dependent variable “dropout”.

0 +β1X1+β2X2+ ……………βnXn

#### **2.3.1 Predictor Variables**

After discussing with the client, we came up with the following list of 16 predictor variables to predict the dropout rate:

1. Year of admission
2. Major/Minor
3. Changed major
4. Courses-related to major
5. High School GPA
6. Transfer—GPA at time of application
7. Grades quarterly
8. Overall Grades
9. Gender
10. Ethnicity
11. Commuting Distance
12. Financial Aid/Scholarship
13. College Credits while in H.S--running start, partnership program where student can earn college credits
14. Age of the student at the time of admission
15. First Generation Group
16. Graduated – Yes/No

**Tools and Techniques**

MS SQL Server, Azure ML Studio, Power BI, Tableau

**2.3.2 Plan ahead**

The plan is to determine the changes that have contributed to a certain student leaving the university and what could be done to remain a student at UW Tacoma. The best method to tackle would be running regression analysis. Additionally, we will discover what motivating variables, if any, can influence the students decision to continue their studies for this we might incorporate using Power BI and SQL. The next steps of the analysis would be:

* Perform correlation analysis
* Split the data to training and validation
* Decide the important features to determine the dropout
* Use Azure ML to run the regression and deploy the model.

Additionally we can use ArcGIS in Power BI to conduct the analysis of commuting distance to the UW Tacoma campus.

**3. A Framework for Agile Analytics**

**3.1 Perform Business Discovery**

***What are we trying to achieve?***

Our goal is to create a predictive model that can determine when students at UWT are at risk for dropping out. The system would serve as a tool for university officials to better understand the increasing number of dropouts and intervene before students take this step.

***What behavior(s) are we trying to understand, influence and/or predict?***

We are attempting to identify academic precursors that manifest before a student drops out of university, to offer targeted resources in an attempt to help persuade the student to complete their degree at UWT

***What type of data would we need to address these issues?***

The data our team has delved into has included:

1) Major

2) Exposure to higher education

3) Engagement in extracurricular activities on campus

4) GPA

5) Years enrolled in the university.

***Is the project even viable?***  
While complicated, the tool has potential to be a groundbreaking tool for universities.

***Does our organization possess the time, budget, and resources to undertake the project?***UWT has dedicated a team for the sole purpose of doing research on student engagement and attributes, there is already the budget and resources collecting the necessary data needed to conduct this analysis.

***Is our organization committed to the project? Or will the project fade into the background as more important priorities crop up?***Yes, the organization is committed and has already dedicated to providing weekly allotted time to meeting with our team as well as providing any extra resources that we may potentially need to complete our project.

***What happens if we don’t answer these questions? What if the project takes longer than  
 expected?***

If we are unable to complete our goal, our analysis will at least provide extra insight to better aid the institutional research team in understanding why graduation rates have been in steady decline over the past eight years.

If the project takes longer than anticipated, we will discuss with the client whether they desire to continue the project further or if our work has been sufficient for them to take over.

**3.2 *Where does the desired data “live?***

The data comes from a number of sources. Student data is often collected every quarter at enrollment, as well as annually by the federal government in financial aid questionnaires.

***Is it even available?***

Partial data is publicly available on the UWT website, as well as on the Integrated Postsecondary Education Data System website. The remainder of the data is stored on a secure database only available for university staff.  
  
 ***Is it legal to use? Is it free to use?***

The public databases are both public and legal to use for anyone who wishes to view the data. As for the secured data, it requires access and proper training in FERPA in order to view. The secured data is not legal to the public if it contains any identifying fields such as address, or name.

***Are we able to retrieve the data in a clean and usable format? Or do we need to scrape it***  
 using one of the tools mentioned in Chapter 2?

Yes, the data is stored in a SQL database and is mostly clean besides for missing data fields.  
  
***Is use of the data restricted?***

The SQL database is a secured database that can only be accessed with permission of the UWTIR staff, in addition, our team is only permitted to access the database with two logins at a time.  
  
***Can you pay to circumvent those restrictions? How much?.***

No, you cannot pay to circumvent any of these restrictions.  
***How long will it take to access/acquire the data?***  
 The public data can be accessed at any time, while the secured data server requires FERPA training to be completed before being granted access to the data.

***How old is our data? Has it aged well?***  
 The database began collecting data in 2016, while there is also public data dating back to the opening of the institution. The data has aged well and has been consistent in collection.  
 ***If the data exists inside of the enterprise, which organizations and departments own the data?***

The data collected by the University itself is collected by the registrar and institutional research team. These are the two organizations dedicated to collecting, researching, and protecting data.

***Is the data complete, accurate and deduplicated?***

Yes, as there is a dedicated team to collecting, cleaning, and researching the data.

**3.3 *Who or what generates the data?***

The students incoming and currently studying at the university generate all the data. While the registrar and UWTIR create questionnaires for the students to help facilitate the data.

***If people are responsible for generating the data, are they trained in how to enter it properly?***

The data is based on academic information on students as well as online questionnaires that can be easily verified.

***Was there turnover in the organization that could introduce inconsistencies and errors? inconsistencies in our data?***

No, the individuals working within the team have stated to be working for several years for UWT.

***Is the data coming directly from the system of record or from another source, such as a datamark or data warehouse?***

The data comes directly from the system of record.

***How is the data currently generated and has that ever changed?***

The data, for the past decade, has been collected online through the registration database as well as the financial aid office.

***How much data is generated?***

As much data as there is students attending the university in addition to past students who have been enrolled within the university.

***What if the data is flawed or incomplete? What are the downsides?***

The downside is students may not be totally honest when answering questions, or there may not be direct correlations in the data we see as potentially useful for our research project.

***Is certain data absolutely required to proceed? What types of proxies can we use if we are left with no choice?***

Approximate data may help us conduct our research, but accuracy will help us eliminate false positives from our algorithm.

***How complex is the data?***

The data is quite complex, as there are many attributes and fields that need to be researched and it can be difficult to find commonalities that are desired.

***How frequently is the data updated?***

The data is constantly being updated and collected, some data can be updated as frequently as once a month, while some can only be collected annually.

3.5 **Score and Deploy**

The UWT Institutional Research project will be deployed utilizing a single model for the university. This model represents the interest for the current student body as well as incoming freshmen and transfer students.

Scoring:

* What is the main goal?
  + To understand what experiences/variables are impacting a student's decision to drop out.
* How will we know we are looking at the right variables?
  + Our project team found 12 distinct variables that are relevant to our model. The variables will help define which students are more likely to drop out. We will enter the evaluation stage where we can observe the tool and acquire data on how effective the tool is. Using the feedback from the evaluation stage, we will determine if the variables selected accurately determine the students from our main goal. If needed, we can go back and select new variables or add additional variables to help further define our goal.
* What weights are you expecting to provide for the variables?
  + Currently, our variables have a generic weight assigned to them based on our initial assessment on the data. Overtime, we can adjust the weights for the variables based on how well the model is determining which students are at risk for dropping out.
* How will we know the model is working well?
  + At this stage of the project, it would be difficult to determine that our model is working well. Working with our client, we will be able to incorporate feedback into our model. This will further define our model and get our project model to a stage where it accurately describes at risk students.

Deployment

* How will the model be deployed?
  + This model will be deployed on an intranet only accessible by UWT’s faculty and staff. Using this tool, staff and faculty will be able to determine what students are at risk and take appropriate action. This could include reaching out to the student and assisting where necessary.
* How confident are you in the predictions?
  + We are working on fully defining what variables would have a substantial effect on our model. We are confident that current variables could predict if a student is at risk but we would like to further research if any additional variables are needed. We would like to incorporate client feedback into our model to help define it further.
* What is our acceptance level of uncertainty?
  + We do not have a current uncertainty level for our model. We will be able to better assess our acceptance level of uncertainty when further modeling has been performed.

**3.6 Evaluate and Improve**

**What data sources are you missing? Which ones are worth including?**

We are missing the Clubs associated, Student involvement, Employment status , Financial aid information columns to further base our analysis and answer our assumptions.

**Which data sources may dry up? What will you do if that happens? (It’s a mistake to assume that a source will be freely available forever just because this is the case today.)**

We are still collecting relevant tables and connecting them in SQL through our remote desktop. However, our access is limited to 2 individuals at the same time.

**Will any data sources be retired?**

No data sources are currently being considered for retirement. All the data that is currently collected is being used and will be updated periodically by UWT IR as needed to maintain the data integrity of the models.

**Will your model improve or diminish if you make fundamental changes?**

The precision of the models should only improve as new and updated data is added. We foresee no reason why the model’s capabilities should diminish at this time.

**What are the time implications?**

As time advances the data sources used for models will need to be updated to reflect changes in variables or new variables ( Student Involvement, Clubs Participated, Survey results) for the qualitative model and new graduation rate, financial aid eligibility for the quantitative model. The client’s timetable is currently forward looking for the next 2 to 4 years meaning that the immediate necessity is not critical and the effect of the models will be seen over time.

**What happens if you make a big mistake?**

The primary area where our models could cause issues for Institutional Research might be incorrect prediction of regression models. A model which does not adequately predict actual dropout with the stopout and missing variables which actually have higher weight than the other features of the model could limit the client’s expectations. The overall effect would be that the University of Washington Tacoma would not be gaining as clear a picture of predicted dropout students profile.

**4. Conclusion**

In conclusion, our data analysis plan addresses the concerning student dropout rate at UWT. Our plan is a comprehensive process that combines descriptive, diagnostic, and predictive analytics.Through analysis of historical data and exploration of different factors such as Pell Grant eligibility, cohort size, majors, and student demographics, our team aims to understand what variables impact student dropout rates. Our findings reveal that variables such as income and prior education experience have potential impact on dropout rates. Our predictive analytics phase introduces a statistical model using Azure ML and logistic regression, leveraging 16 predictor variables to forecast dropout likelihood. The deployment plan is focused on making the tool accessible to UWT’s faculty and staff.

Our agile analytics framework emphasizes a continuous understanding of our main goal and data sources. The score and deploy phase’s focal point is further refining our model through client feedback and adjusting our variables/variable weight for improved accuracy.

Despite our current limitations such as missing data sources and potential future data changes, our partnership with UWT and their research team is a great combination for success. As the model evolves, we are expecting to see valuable insights into the factors that impact student dropout. Our proactive approach outlined in this plan overall focuses on figuring out why there is a growing student dropout rate but also how the university can take timely and effective measures to enhance student retention.