PURDUE STUDENT LIFE & RECWELL

Tyler Glass, Kyle Stowe, Anushka Gupta, Elsa Jerry, and Matt Hui

INTRODUCTION

Mission Statements:

Student Life - "Student life at Purdue fosters student success through learning opportunities inside and outside of the classroom. We focus on co-curricular education and leadership development as integral parts of the Purdue experience."

Purdue Recreation & Wellness - "Provide the Purdue community with recreational and wellness activities that contribute to learning and pursing an active, healthy lifestyle."

GOAL

- Analyze both academic and non-academic variables that might correlate to student's academic and mental success within Purdue's student life
- Demonstrate variables within Purdue Recreational and Wellness that might affect student's GPA through the use of dashboard on Tableau

ACKNOWLEDGMENTS

- Rishabh Rajesh – Team Mentor & TA
- Carl Krieger – Director of Residential Life
- Abigail Vorhies – Assistant Director at RecWell Sport Program
- Mike Seals – Associate Director, Strategic Initiatives & Assessment

TECHNOLOGY USED

- R Programing Language
- Python Programing Language
- Tableau

CONCLUSION

- We worked with PCA and K means a lot this semester and using these strategies, you can find out a lot of information that you couldn't find using other methods.
- We were able to find a lot of GPA data using 2d and 3d graphs that we made using these methods and we found more data with the recwell using the same methods.

METHODS

Purdue Recreation & Wellness:

In this research, we focused on two different projects/sides: Purdue Student Life and Purdue Recreation & Wellness.

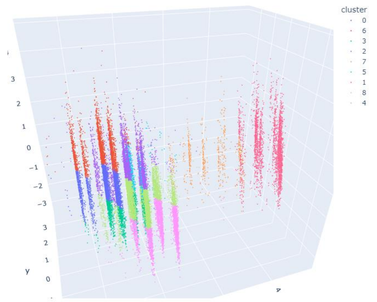
Purdue Student Life:

- **Cluster initialized** – 9
- **Variables used** –
  - Term GPA, Gender, Residency, Standardized Test Z-scores, Class, Degree Classification, Credit Hierarchy, On/Off campus, and degree code
- **Steps** –
  - Clean all the data in R
    - Use labelencoder to transform non-numerical labels to numerical labels
  - Using PCA to narrow the multidimensional dataset into two or three dimensions
  - Using K-means analysis to group the 2 or 3-dimensional data in 9 different cluster/groups

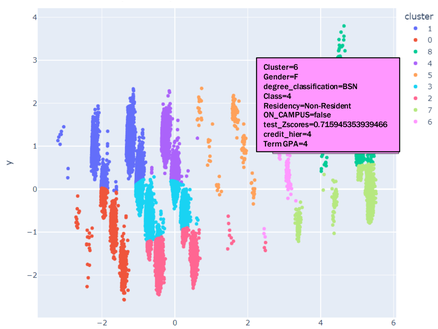
Purdue Recreation & Wellness:

- **Cluster initialized** – 8
- **Variables used** –
  - Granted, Gender, Residency, Ethnicity, Residence Hall, Program, Classification, Term GPA, Credit Hierarchy, and Credit Dropped
- **Steps** –
  - Preprocess all the data in Python
    - Use labelencoder to transform non-numerical labels to numerical labels
  - Using PCA to narrow the multidimensional dataset into two or three dimensions
  - Using K-means analysis to group the 2 or 3-dimensional data in 9 different cluster/groups

PCA K-means | 3 Components and 9 Clusters



PCA K-means | 2 Components and 9 Clusters



CLUSTER ANALYSIS

Cluster 4:
- Majority consists of BSE (Bachelors of Engineering) degree type
- Higher standardized test scores
- Lower average term GPA

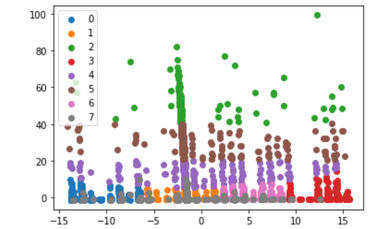
Cluster 6:
- Majority consists of BSN (Bachelors of Nursing) degree type
- Female dominated cluster
- Higher average term GPA

STUDENT LIFE ANALYSIS

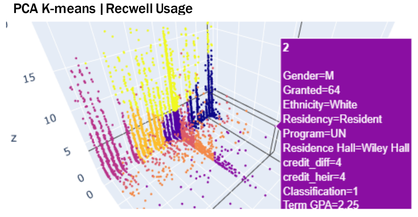
2-Dimensional PCA & K-means:
- There is a variance of 64.3% that was captured through PCA
  Students that are not part of the condensed areas of each of the clusters tend to have a GPA of 0
- Each of the nine columns created in the PCA K-means graph could suggest the nine different degree types
- Within each of the nine columns created, there are gaps creating four different groups, which is likely to be indicating each of the four classes (Freshman, Sophomore, Junior and Senior)

3-Dimensional PCA & K-means:
- There is a variance of 77.81% that was captured through PCA
- Through the 3-dimensional view, one can see multiple columns behind each other
- Due to the large quantity of chaos that the PCA K-means created, we removed test score from the 3-dimensional graph
  - The chaos is likely caused by the sheer range and variance of test scores across the dataset

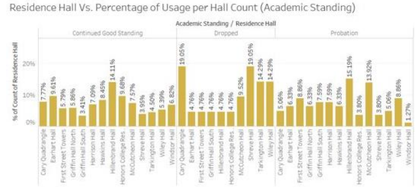
PCA K-means | 2 Components and 9 Clusters



PCA K-means | Recwell Usage



Residence Hall Vs. Percentage of Usage per Hall Count (Academic Standing)



RECREATION & WELLNESS ANALYSIS

Problem:
- We found that the data was too stretched out to find anything conclusive.

Some things we can find:
- We were able to find a few different things on the largely populated streak in the middle
- This tall streak is students that live off campus
- What we learned about students living off campus is that they tend to use the corec facilities more
- Even though they use the corec more, they tend to also drop classes more often than students living on campus.

FUTURE PROJECTS AND PLANS

- **Adding non-academic factors to the analysis**
- **Adding retention rates and social engagement as an additional student success measurement**
- **Apply our PCA K-means method with student dining hall data**
- **Analyze student data across different years to identify a student's progression**
- **Using Tableau dashboard to help visualize our machine learning results**