

# ECON-562: Project 1

---

US Dept of Education  
College Scorecard  
Andrew Frohner

John Garcia  
March 24<sup>th</sup>, 2025

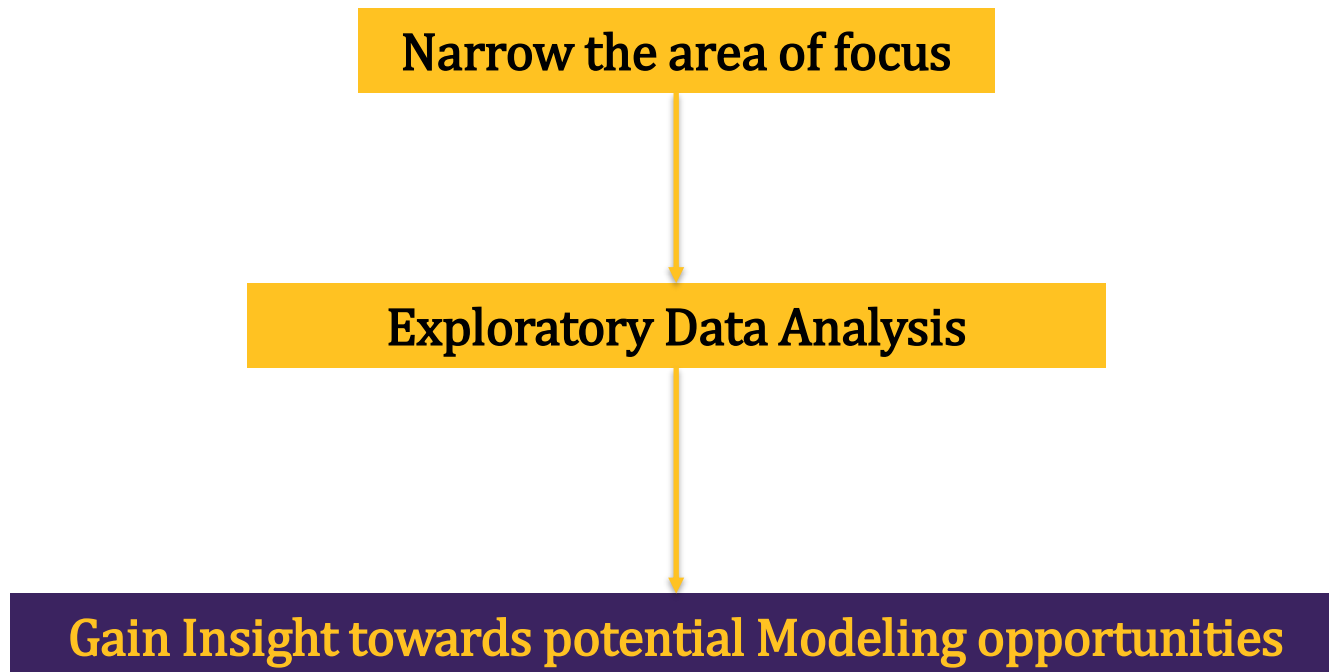
California Lutheran University

---

**Master of Science in Quantitative Economics**

# Purpose of Analysis

---



# Data Review - Overall

---

## Data Home

Download the data that appear on the College Scorecard, as well as supporting data on student completion, debt and repayment, earnings, and more.


This data was **last updated January 16, 2025**.

### All Data Files


[Download \(.zip, 390 MB\)](#) 

- Institution-level data files for 1996-97 through 2022-23 containing aggregate data for each institution. Includes information on institutional characteristics, enrollment, student aid, costs, and student outcomes.
- Field of study-level data files for the pooled 2014-15, 2015-16 award years through the pooled 2018-19, 2019-20 award years containing data at the credential level and 4-digit CIP code combination for each institution. Includes information on cumulative debt at graduation and earnings one year after graduation.
- Crosswalk files for 2000-01 through 2022-23 that link the Department's OPEID with an IPEDS UNITID for each institution.

### Most Recent Institution-Level Data

[Download \(.zip, 22 MB\)](#) 

### Most Recent Data by Field of Study

[Download \(.zip, 13 MB\)](#) 

**Source:** US Dept of Education

**Year:** 2022 – 2023

**Number of Observations:** 6484 schools

**Number of Dimensions:** Over 3,200 individual features

# My Data

---

## Features explored:

- Demographic factors (Segregated and Co-ed universities)
- Financial Aid (Pell Grant, Federal loan %s)
- Student Outcomes (completion rate)
- Student Financial Outcomes (default rate , completion rate)
- Institutional factors (size, 2 yr or 4 yr degree, Public/Private, Cost of attendance)

# Exploratory Data – Focus

---

## Student Outcomes

# Exploratory Data – Curiosity

---

## Student Default Rate

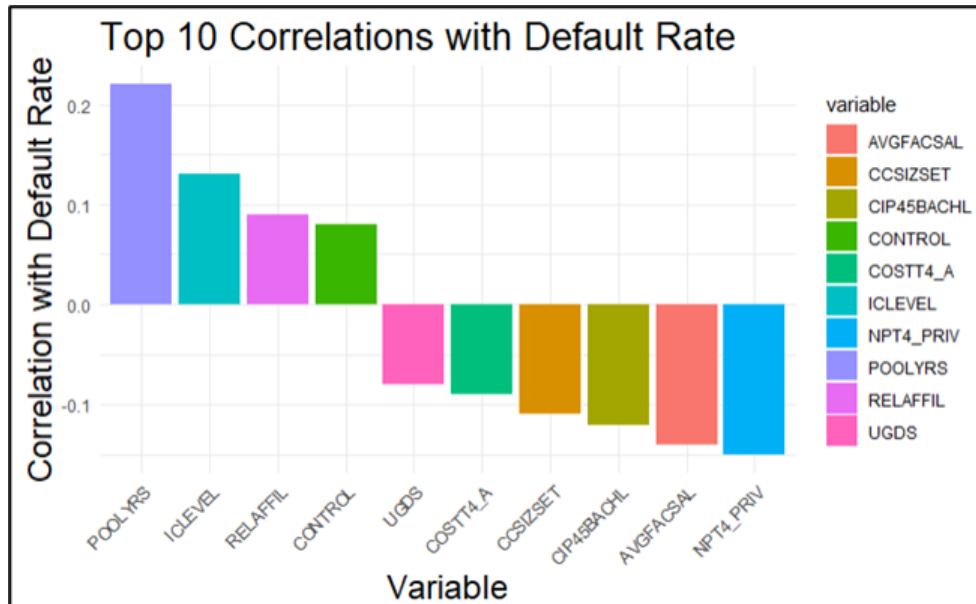
**Research question:** Can other predictors be used to categorize student default rates?

## Student Completion Rate

**Research question:** What is related to student default rate at large schools

# Exploratory Data – Analysis

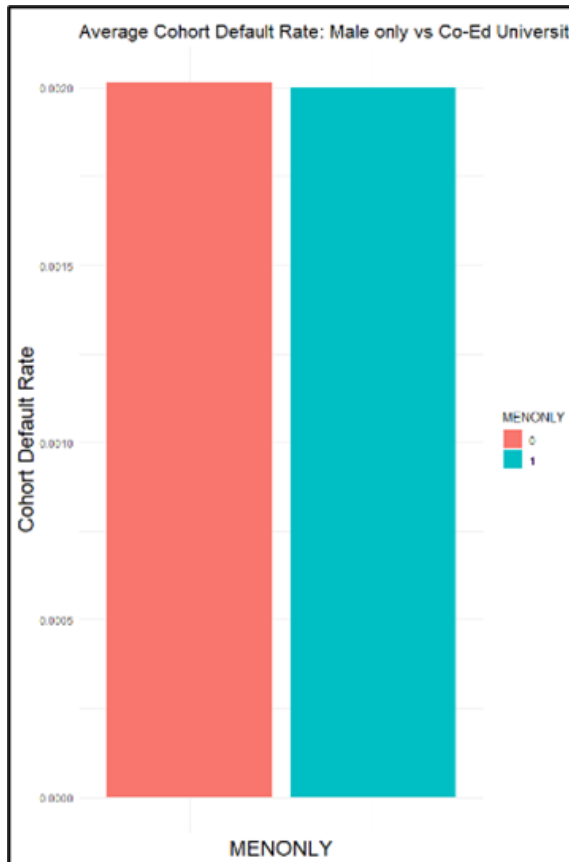
Default Rate does not exhibit strong correlation with other predictors



# Exploratory Data – Analysis

---

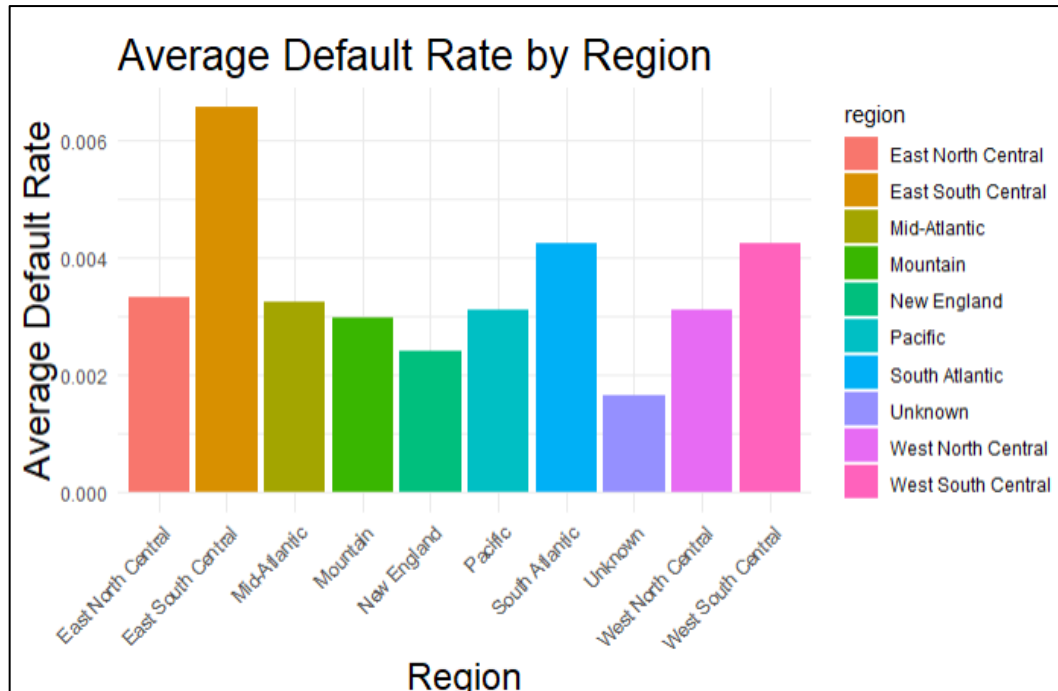
Default Rate did not appear to be higher for Men only schools





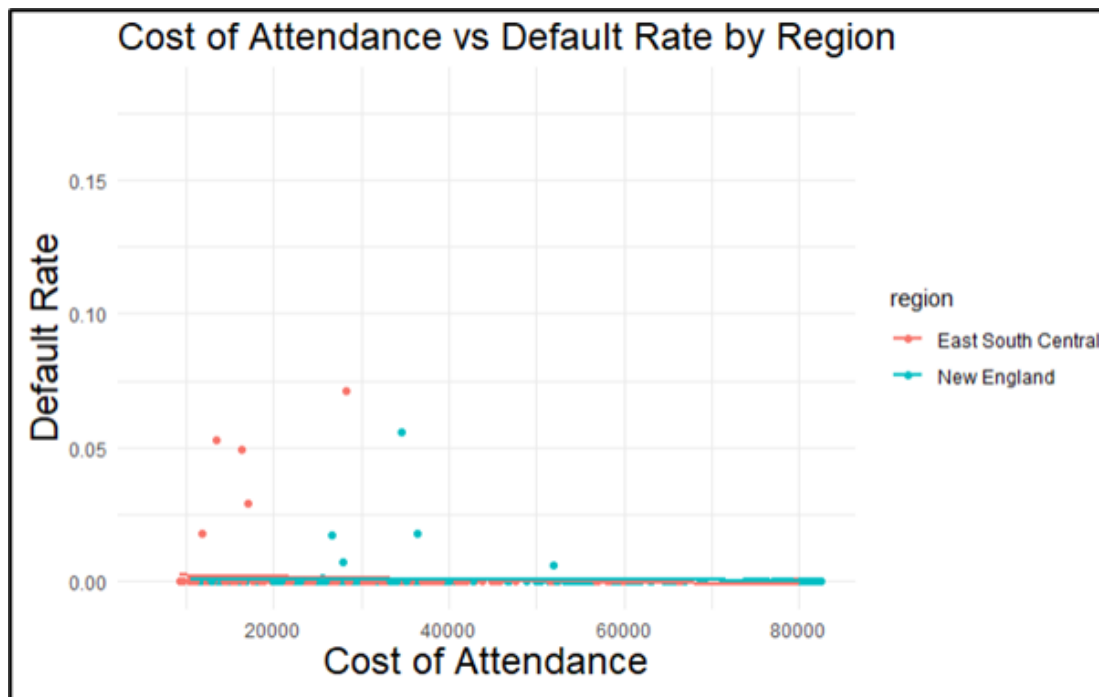
# Exploratory Data – Analysis

Default Rate appeared to be mostly stable across US Census Regions



# Exploratory Data – Analysis

Default rate relationship with a correlated variable across High and Low default regions



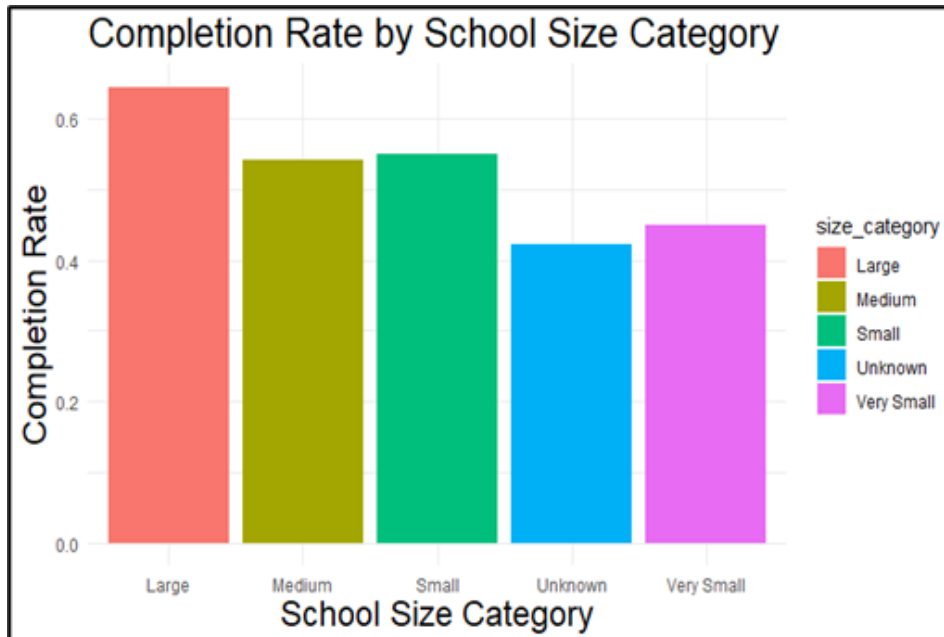
---

**Slide intentionally left blank**

# Exploratory Data – Analysis

---

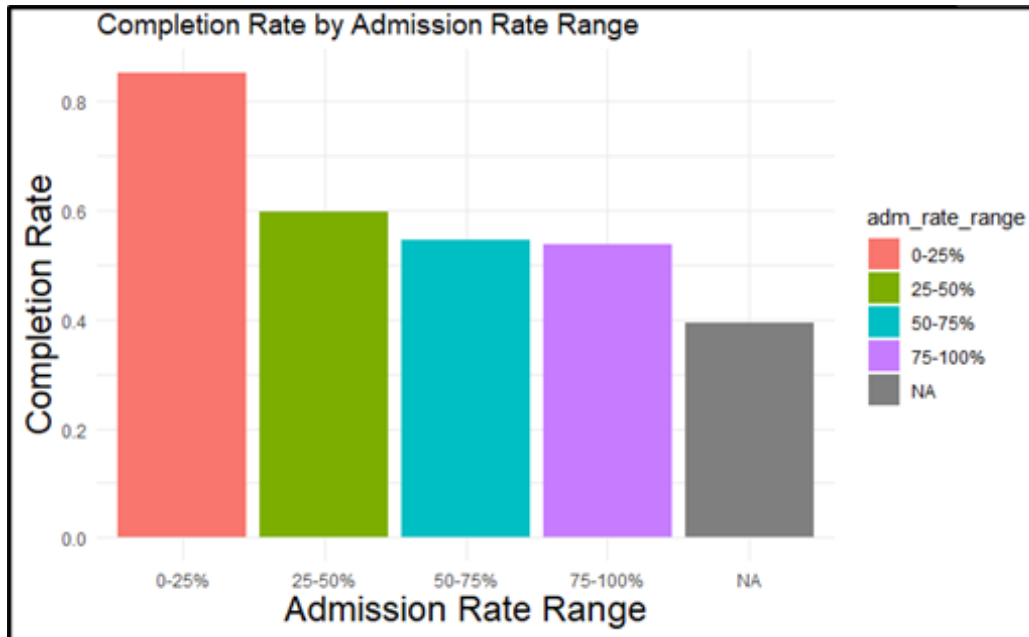
Completion Rate steps down as the university gets smaller...



# Exploratory Data – Analysis

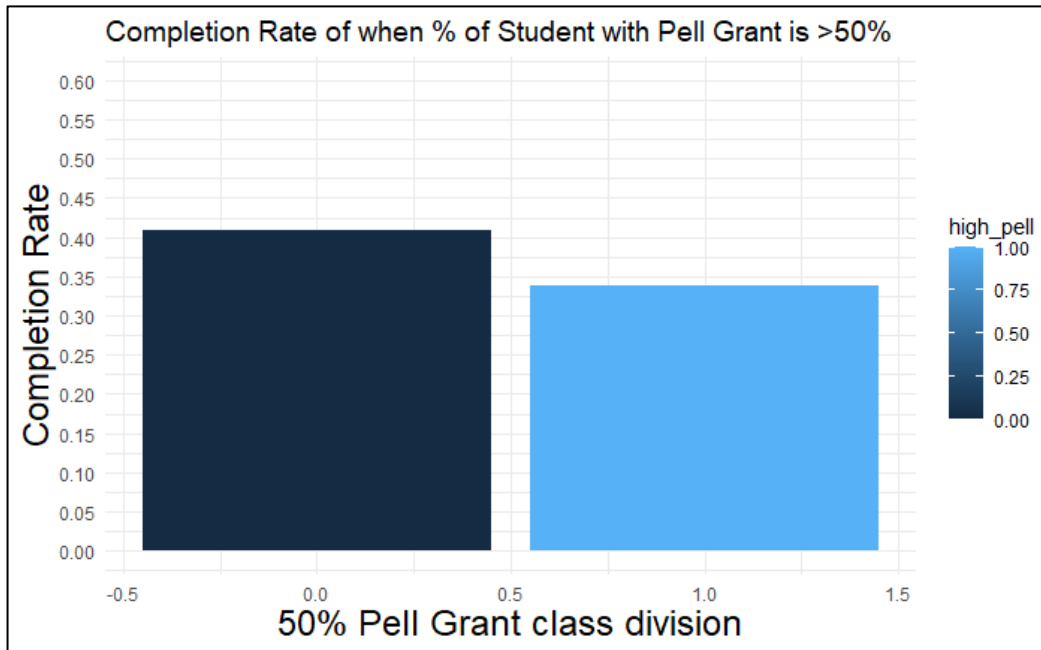
---

Completion Rate steps down as the admission rate gets higher



# Exploratory Data – Analysis

Universities with large population of Pell grant students have lower completion rates



# Takeaways - Summary

---

## 1. Default Rate

- Prediction or classification will require further EDA
  - Direct focus on student cost variables/University cost variables

## 2. Student Completion Rate

- Clear decreasing relationship with Institutional factors
- Surprising relationship with Financial Aid

---

Thanks !



# Exploratory Data – Analysis

## Institutional Factors by Default Rate categories

