Project Presentation Outline

1. Objective of our project
   1. The summer before students’ senior year at high schools, it is usually recommended that parents and students start to do the preliminary work for college admission. This includes visiting colleges to narrow down the choices. Most students/parents narrow down their choices to 5 schools.
   2. The Prince George's County Board of Education has asked our data team to make a presentation for junior year high school students and parents to help them in their decision on what colleges or universities they should consider. Given the cost of tuition, commutable distance, and the number of highly rated colleges and universities, we decided to focus on colleges and universities within the Mid East Region and Virginia. The presentation will detail various data supported advantages and disadvantages of matriculating to higher education institutions within the Mid East Region and Virginia.
   3. Our “Research Question” is: Which universities or colleges should 11th grade students from Prince George’s County public schools consider attending, and what characteristics should they take into consideration with respect to academic student success and post-graduation financial stability?
2. Overview of the data set
   1. The dataset that our group has chosen to use for our team project is the US Department of Education’s College Scorecard dataset, which is freely accessible online from the department’s Scorecard website (https://collegescorecard.ed.gov/data/). Specifically, this data contains records on all Title IV aid receiving higher education institutions in the US, as indicated by the Integrated Postsecondary Education Data System (IPEDS) federal reporting surveys. The purpose of the US Department of Education providing these data freely online is so that prospective high school students and their parents could have a way to compare US higher education institutions across a wide range of data points dealing with college costs and student success outcomes, in order to allow them to make the most data driven and informed opinion as possible on where potential institutions to attend. The data are compiled by the US Department of Education using federal reporting from institutions, data on federal financial aid, and tax information.

* 1. There are several data files made available for consumers, but we have chosen to use the “Most recent data” dataset provided on the College Scorecards main website that was linked in the previous paragraph, but the exact URL to download our chosen data is (https://ed-public-download.app.cloud.gov/downloads/Most-Recent-Cohorts-All-Data-Elements.csv).

1. Data Preparation
   1. We filtered the data by
      1. States - [DE, DC, MD, NJ, NY, PA, VA (ST\_FIPS = 10,11,24,34,42,51)]
      2. Institution awarding – Associate Degree, Bachelor Degree, Graduate Degree (HIGHDEG = 2,3,4)
      3. Institutions that are currently operating – (CURROPER = 1)
      4. Exclude schools that are on heightened cash monitoring by the Department of Education (HCM2 = 0)
      5. Exclude institution that are exclusively for graduates (CCSIZET not equal 18)
      6. Added descriptive columns for 8 columns
         1. Institutions under investigation - HCM2,
         2. Predominant Undergrad degree awarded - (PRED\_DEG\_DESC),
         3. Highest Degree Awarded - (HIGH\_DEG\_DESC),
         4. Public or Private Institution - (CONT\_DESC),
         5. State Code (ST\_FIPS\_DESC),
         6. Region (REGION\_DESC),
         7. Locale (LOCALE\_DESC),
         8. Carnegie Classifications (CCBASIC\_DESC, CCUGPROF\_DESC, CCSIZE\_DESC)
      7. The original data set had 1,779 variables. We reduced it to 150 variables
2. Identification and analysis of the key factors in selecting a higher learning institution
   1. There are 4 key factors that a parent and/or student uses to evaluate a college or university
      1. Financial stability of the institution
      2. Academic success of the institution
      3. Tuition cost – provided by the data set
      4. Proximity of the institution to home – provided by the data set
   2. To calculate the financial stability of the institutions within the region, we will calculate the z-score for each of the following variable. We will then add the z-scores to arrive at the financial stability super score
      1. Default Rate (2 yrs & 3yrs)
      2. Average student earnings 10 year after graduating
      3. Median debt of graduates
      4. In-State Tuition
      5. Out-Of-State Tuition
   3. To calculate the academic success of an institution within the region, we will calculate the z-score for each of the following variables. We will then add the z-scores to arrive at the academic success super score
      1. First-time full-time student retention rate
      2. First-time, full-time student completion rate
      3. Percentage withdrawn from original institution within 3 years
   4. We will then demonstrate how these factors is to identify the top 10 schools
   5. We will then demonstrate for 4 different profiles, how they can apply these factors to select their top 5 schools
3. Super Score Data Set
   1. To create the SuperScore Data, we created a subset with the following variables
      1. Institution ID (OPEID)
      2. State - (ST\_FIPS\_DESC)
      3. local of institution - (LOCALE\_DESC)
      4. public or private institution – (CONT\_DESC)
      5. Carnegie rating – (CCSIZE\_DESC)
      6. tuition costs – (TUITIONFEE\_IN, TUITIONFEE\_OUT)
      7. proportion of faculty staff full-time (PFTFAC)
      8. completion rate – (C150\_4\_C150\_L4)
      9. retention rates – (RET\_FT4\_FTL4)
      10. default rates. – (CDR2, CDR3)
      11. average student earnings – (MN\_EARN\_WNE\_P10)
      12. median loan debt of graduates – (GRAD\_DEBT\_MDN\_SUPP)
      13. withdrawal rate – (WDRAW\_ORIG\_YR3\_RT)
4. Descriptive Analysis of the SuperScore Data Set
   1. Number of colleges and universities in each state in the Region
   2. Number of Public, Private Non-Profit, and Private for profit colleges and universities in the Region
   3. Top 10 schools with proportion of faculty full-time
   4. Range of Default Rate within the Region and by the States within the Region
   5. Range of Average Student Earnings within the Region and by the States within the Region
   6. Range of Median Debt within the Region and by States within the Region
   7. Range of In-State Tuition for the Region and by States within the Region
   8. Range of Out-of-State Tuition for the Region and by States within the Region
   9. Range of Retention Rate within the Region and by States within the Region
   10. Range of Completion Rate within the Region and by the States within the Region
   11. Range of Withdrawal Rate within the Region and by the States within the Region
   12. 10 schools with lowest default (2&3yrs)
   13. Top 10 schools by average student earning
   14. 10 schools with lowest median debt
   15. 10 schools with the lowest in-state tuition
   16. 10 schools with the lowest out-of-state tuition
   17. Top 10 schools by retention rate
   18. Top 10 schools by completion rate
   19. 10 schools with lowest withdrawal rate
5. Quantitative Analysis of the SuperScore Data Set
   1. Financial Stability
      1. Mean & Standard Deviation of Default Rate (2 yrs & 3yrs)
      2. Mean & Standard Deviation of Average student earnings 10 year after graduating
      3. Mean & Standard Deviation of Median debt of graduates
      4. Mean & Standard Deviation of In-State Tuition
      5. Mean & Standard Deviation of Out-Of-State Tuition
   2. Academic Success
      1. Mean & Standard Deviation of First-time full-time student retention rate
      2. Mean & Standard Deviation of First-time, full-time student completion rate
      3. Mean & Standard Deviation of Percentage withdrawn from original institution within 3 years
   3. Calculate Z-Score for Financial Stability variables
   4. Calculate z-score for Academic Success variables
   5. Calculate Financial Stability SuperScore for each institution
   6. Calculate Academic Success SuperScore for each institution
6. Demonstration of how the 4 key factors can be used to identify top 10 institutions
   1. Top 10 institutions within the region
   2. Top 10 institutions within a state within the region
7. Demonstrate for 4 different profiles,
   1. Apply key factors to select top 5 schools for Profile A
   2. Apply key factors to select top 5 schools for Profile B
   3. Apply key factors to select top 5 schools for Profile C
   4. Apply key factors to select top 5 schools for Profile D