Project Presentation Outline

1. Project Overview
   1. 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.
   2. 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 how we prepare and analyzed the data and create a College Selection Model that students and parents can use to select the top 5 colleges that matches their profile.
2. Research Question
   1. 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
3. 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. Our Approach for creating a College Selection Model
   1. There are 3 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. Affordability
   2. To calculate the financial stability of the institutions within the region, 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. 2 yrs. Default Rate
      2. 3 yrs. Default Rate
      3. Average student earnings 10 year after graduating
      4. Median debt of graduates
   3. To calculate the academic success of an institution within the region, 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. To calculate the affordability of attending 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 affordability super score
      1. In-State Tuition
      2. Out-of-State Tuition
   5. Add the z-scores and superscores to our original data set
   6. Create a subset with the selection criteria that we are expecting the parent/student to provide
      1. Income Level
      2. Percentile of SAT
      3. Percentile of ACT
      4. Type of Degree (Associate or Bachelor)
      5. School Preference (Private or Local)
      6. State (DC, MD, DE, PA, NJ, NY, VA)
      7. Local (City, Urban, Town, or Rural)
   7. Create 4 profiles, and use the College Selection Model to identify the top 5 schools for each profile
3. Create Superscore Data Set
   1. To create the Superscore Data Set, 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. Barplot the number of colleges and universities by state in the Region
   2. Barplot of the number of Public, Private Non-Profit, and Private for profit colleges and universities in the Region
   3. Barplot the mean of the default rate by States within the Region
   4. Barplot of the mean of the average student earnings by States within the Region
   5. Barplot the mean of the Median Debt by States within the Region
   6. Barplot the mean of the In-State Tuition by States within the Region
   7. Barplot the mean of the Out-of-State Tuition by States within the Region
   8. Barplot the mean of the Retention Rate by States within the Region
   9. Barplot the mean of the Completion Rate by the States within the Region
   10. Barplot the mean of the Withdrawal Rate by the States within the Region
   11. Top 10 schools by proportion of faculty full-time
   12. 10 schools with lowest default (2&3yrs)
   13. Top 10 schools with the highest 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 with the highest retention rate
   18. Top 10 schools with the highest completion rate
   19. Top 10 schools with the lowest withdrawal rate
5. Quantitative Analysis of the Superscore Data Set
   1. Financial Stability
      1. Mean, Standard Deviation & Histogram of Default Rate (2 yrs & 3yrs)
      2. Mean, Standard Deviation & Histogram of Average student earnings 10 year after graduating
      3. Mean, Standard Deviation & Histogram of Median debt of graduates
   2. Academic Success
      1. Mean Standard Deviation & Histogram of First-time full-time student retention rate
      2. Mean, Standard Deviation & Histogram of First-time, full-time student completion rate
      3. Mean, Standard Deviation & Histogram of Percentage withdrawn from original institution within 3 years
   3. Affordability
      1. Mean, Standard Deviation & Histogram of In-State Tuition
      2. Mean, Standard Deviation & Histogram of Out-Of-State Tuition
   4. Calculate Z-Score for Financial Stability variables
   5. Calculate z-score for Academic Success variables
   6. Calculate z-score for Affordability variables
   7. Calculate Financial Stability Superscore
   8. Calculate Academic Success Superscore
   9. Calculate Affordability Superscore
6. Create the College Selection Model with the following variables
   1. Institution ID (OPEID)
   2. Institution Name (INSTNM)
   3. State - (ST\_FIPS\_DESC)
   4. local of institution - (LOCALE\_DESC)
   5. public or private institution – (CONT\_DESC)
   6. Carnegie rating – (CCSIZE\_DESC)
   7. tuition costs – (TUITIONFEE\_IN, TUITIONFEE\_OUT)
   8. proportion of faculty staff full-time (PFTFAC)
   9. completion rate – (C150\_4\_C150\_L4)
   10. retention rates – (RET\_FT4\_FTL4)
   11. default rates. – (CDR2, CDR3)
   12. average student earnings – (MN\_EARN\_WNE\_P10)
   13. median loan debt of graduates – (GRAD\_DEBT\_MDN\_SUPP)
   14. withdrawal rate – (WDRAW\_ORIG\_YR3\_RT)
   15. Average net price by income level – public (NPT41\_PUB, NPT42\_PUB, NPT43\_PUB, NPT44\_PUB, NPT45\_PUB)
   16. Average net price by income level – private (NPT41\_PRIV, NPT42\_PRIV, NPT43\_PRIV, NPT44\_PRIV, NPT45\_PRIV)
   17. 25 Percentile of SAT (SATVR25, SATMT25, SATWR25)
   18. Midpoint of SAT (SATVRMID, SATMTMID, SATWRMID)
   19. 75 Percentile of SAT (SATVR75, SATMT75, SATWR75)
   20. 25 Percentile of ACT (ACTEN25, ACTMT25, ACTWR25)
   21. Midpoint of ACT (ACTENMID, ACTMTMID, ACTWRMID)
   22. 75 Percentile of ACT (ACTEN75, ACTMT75, ACTWR75)
   23. Type of Degree (HIGH\_DEG\_DESC)
   24. Z-score columns
   25. Superscore columns
7. Evaluate the College Selection Module with 4 personas
   1. Select top 5 schools for Persona A
   2. Select top 5 schools for Persona B
   3. Select top 5 schools for Persona C
   4. Select top 5 schools for Persona D