**USA College Graduated Student Rates**

Data Source:

cc\_institution\_details.csv

cc\_institution\_grads.csv

nst-est2009-01.csv

nst-est2005-01.csv

Description for table cc\_institution\_details.csv:

*This table contained information for 3800 institutions throughout the USA. Data per institution included location (city, state), median SAT score, sector (public or private), and average financial aid granted*.

Field descriptions:

**unitid** - Department of Education Unit ID number

**chronname** - Institution name

**city** - Institution city

**state** - Institution state

**level** - Level of institution (4-year, 2-year)

**control** - Control of institution (Public, Private not-for-profit, Private for-profit)

**student\_count** - Total number of undergraduates in 2010

**med\_sat\_value** - Median estimated SAT value for incoming students

**aid\_value** - The average amount of student aid going to undergraduate recipients

**aid\_percentile** - Institution's percent rank for average amount of student aid going to undergraduate recipients within sector

Description for table cc\_institution\_grads.csv:

*This table contained about 1,302,102 records of groups of students who graduated within 4 years and 6 years by school for the years 2002 to 2013.*

Field descriptions:

**unitid** - Department of Education Unit ID number.

**year** - Year of data release.

**gender** - Gender of students ('B' = both genders; 'M' = male; 'F' = female)

**race** - race/ethnicity of students ('X' = all students; 'Ai' = American Indian; 'A' = Asian; 'B' = Black; 'H' = Hispanic; 'W' = White)

**cohort** - degree-seeking cohort type ('4y bach' = Bachelor's/equivalent-seeking cohort at 4-year institutions; '4y other' = Students seeking another type of degree or certificate at a 4-year institution; '2y all' = Degree-seeking students at 2-year institutions)

**grad\_cohort** - Number of first-time, full-time, degree-seeking students in the cohort being tracked, minus any exclusions

**grad\_100** - Number of students who graduated within 100 percent of normal/expected time

**grad\_150** - Number of students who graduated within 150 percent of normal/expected time

Description for table: nst-est2009-01.csv

*This table contained census estimates based on the year 2000 with projections until the year 2009.*

Description for table nst-est2015-01.csv

*This table contained census estimates based on the year 2010 with projections until the year 2015.*

Data Cleaning:

Due to an incomplete data set with missing values the two main data sets were cleaned and altered to help with the analysis of the data to find our objective.

Modifications to institution\_details (Institutions) and institution\_grads (Grads) data sets were as follow:

1.-Number of fields were reduced to 10 fields in Institution’s data set and 7 fields in Grad’s data set.

2.-year colleges were not included in our findings so we dropped those records in the Institution’s data set to only keep 4-year colleges.

3.-Institutions with rows that had null values were be dropped. By doing that we lost data integrity because the number of Institutions we kept for our study was no longer matching our institutions in the Grads data set.

4.-To correct the above an array of schools from our Grads data set that did not exist in the Institutions data set was created to then erase all those students from our Grads data set to maintain data integrity among the data sets. Making our sample of study much smaller.

5.- Also the distribution of college graduates by gender and the ethnicity had to be dropped because of lack of data. The data also contained the summarized records of all ethnicities and both genders with full data, classified as X and B respectively, making it easier to just pull those records for our analysis.

6.- The data of college graduates was classified by year, state, schools and groups of study or students that were seeking a 4-year degree. We grouped the data by year, state and school summarizing the number of students that graduated within 4 years and within 6 years.

7.-In order to single out the state with more graduates per year and per state and analyze its correlations, we had to take some factors into consideration. One of them was the population.

8.- Population tables were pulled out of the census websites to attach to our datasets so we could calculate the number of graduates per capita that graduated in that year and state.

9.-There were two census tables that needed to be cleaned and manipulated to be merged into our Grads table by state and year.

Findings:

1.- Once we had the census, we were able to single out by year which state gave us the largest number of graduates per capita.

2.- The District of Columbia had largest number of graduates per capita. Not surprising because the population is smaller than other states.

3.-The state with the largest number of graduates without taking into consideration population was California. The data showed that the year 2013 was the year with most graduates.

4.- After singling out our state of California for the year 2013 we will continue with our analysis and find the correlations of SAT score, financial aid and sectors.