# Student Fraud Ring
# Data Analytical Project Summary Guide Points

**Pre-qualifier** – the project I will be discussing is still ongoing and is in the final predictive analytics phase. It also involves sensitive data points surrounding student fraud ring activity within the Post-Secondary arena that I will not be able to obviously disclose. Having said that…per legal counsel ☺ I have selected this project for this discussion because I believe the lessons learned and best practices can be readily understood and mimicked in other data analytic projects.

**Background** – What is termed student fraud rings are a rapidly growing crime that have targeted U.S. Department of Education federal student aid programs, with a potential impact on millions of students at thousands of institutions of higher learning nationwide. The department processes over 19 million electronic applications[1] for student financial aid and disburses over $90 Billion[2] a year in federal student aid funds. The ability for criminals to submit multiple factious applications using either 'straw men' or stolen identity information is a rapidly expanding concern.

**Brain Storming Sessions** - Developed a list of all possible fraud indicators that may identify a student fraud ring. For example:

- Address
- Phone number
- IP address
- Security questions/answers etc..

**Data Source/ Data Cleansing** –

- Identified pilot time frame for snap shot of data.
- Requested from investigations a listing of schools that contain known fraud ring activity.
- Addresses – discovered that address can be in many forms such as 1st/First/Furst St/Street/Stret.
    - Contacted Postal Service and identified a certified vendor who had address filtering system to standardize all addresses.
    - Ran all addresses against this system.
- Acquired Data extracts / Data Warehouse
    - Met with each related ED application manager and reached agreement to acquire specified data snap shot based on defined time line and data elements.
    - Built Oracle Data Warehouse to be able to handle high volume and organize in indexed reciprocal tables.

**Interim Reporting** - Developed Trend and Rule Based Exception Reports using identified fraud indicators to assess the existing data population. Basically this is for reasonableness testing to gain assurance that the data identified such as multiple occurrences of address, emails etc.

---

[1] **Number of distinct applicants' filing FAFSA applications on the CPS system for the 2010 school year. For those applicants awarded, there were 15,937,379 distinct applications in NSLDS who received Title IV (Grants and/or Loans) in School year 2010.**

[2] **Total student loans disbursed in School year 2010 as recorded in the NSLDS system.**

**Risk Model / Data Reduction** - Aggregate or amalgamate the information contained in large datasets into manageable (smaller) information nuggets to filter out false positives.

**Light Bulb Moment**

- Dark before the Light – we delegated the assessment of the risk model results to all of the project team members, but it became readily apparent that there was still 1000's of groups of potential fraud rings to the point that the manual assessment was not doable.
- Back to square one – we went back to the beginning and had an investigator walk thru line by line every aspect of a number of actual successful fraud ring cases. All of sudden we realized that there were key fraud ring indicators that stood out from the others.
- After additional 'white board' analysis, we built a new filtering model that specified three key fraud indicators that required a potential fraud ring grouping to meet.
- Basically a fraud group had to have at least [x] or more common matches of these key fraud indicators.

**Final results**

- Within the selected school pilot population we were able to filter out 99% of the false positives, for example if we had a total population in these schools of 100,000 students, we would have reduced it to 1,000 possible fraud ring students.
- We had identified ALL known fraud rings.
- We had identified new fraud rings.
- We had added students to existing fraud rings under investigation.
- Taking a step back and looking at our current filtering process, it appears obvious but in reality it took some work to get here from there.

**Statistically Supportable** – we now had a very good idea about the possible level and trending of student fraud ring activity nationwide but the data results in this phase of the data analytical process are not statistically supportable. We are now in the process of completing what is called the Unsupervised Reporting phase which incorporates statistically viable processes where we can say with a high degree of probability what and where fraud rings are nationwide.

**Unintended Consequences** – though our projects objective was to identify fraud rings for our investigative staff, we are able to:

- Identify what the big picture most likely is. Up to this point, no one really knows how large this problem is.
- Impact on the Financial Statements – As the primary lender in the student loan arena, financial auditors want to know what is the potential impact on the financial statement?
- Are we working the right cases – Effective resource management of limited investigation assets? Up till now, our cases have been primarily from hot line or school referrals. There may well be very large rings out there that were up to now hidden in the background.