BRIGHT SPOT ANALYSIS

KAVYA BEHERAJ

MARCH 14, 2018

BACKGROUND

STUDENT SUCCESS NETWORK
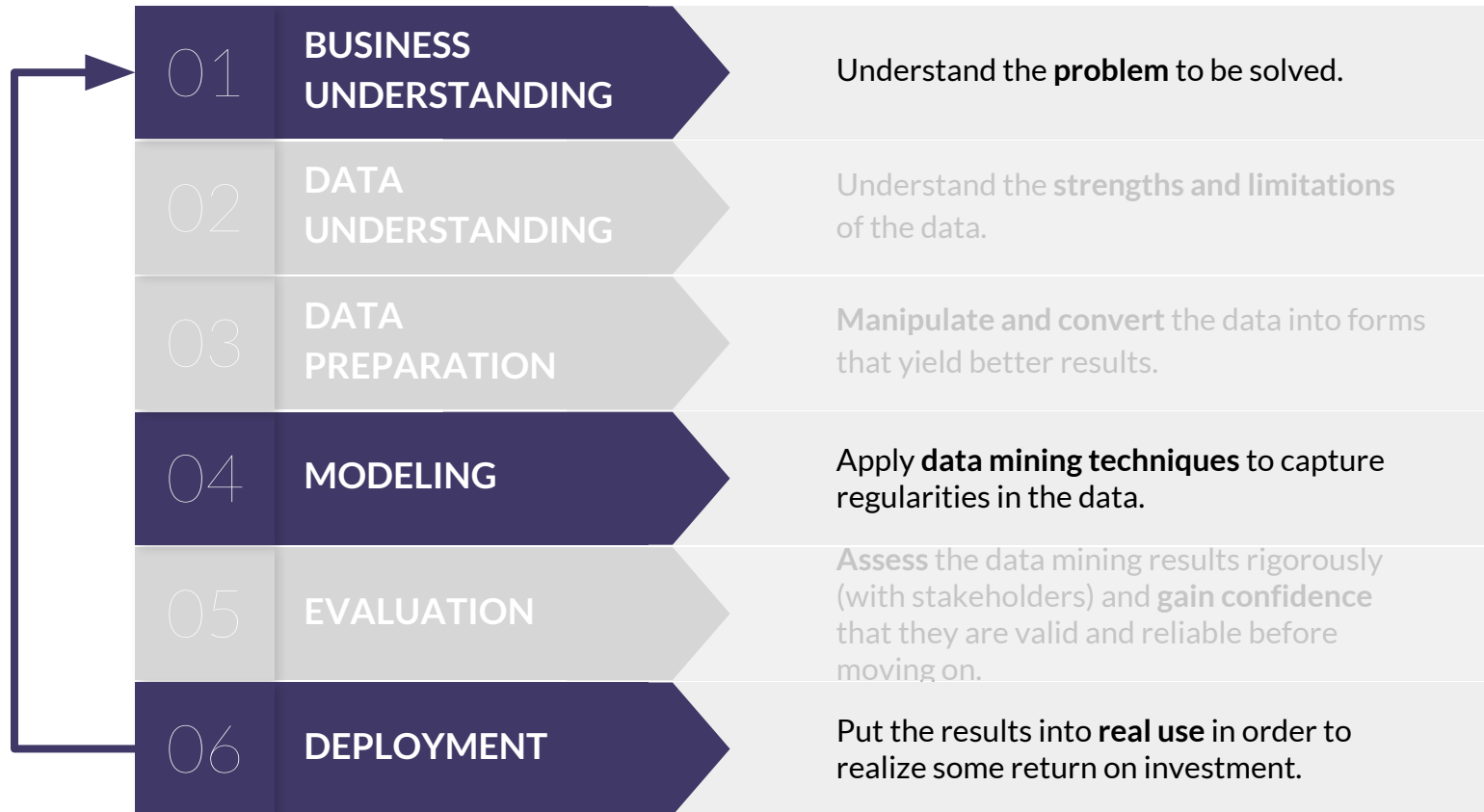
The Research Alliance for New York City Schools

Better Evidence for Better Schools

The Student Success Network is a community of **50 nonprofit organizations** in NYC committed to using data to improve student **social-emotional learning**.

The Research Alliance is a center housed at NYU that conducts **rigorous, applied research** on NYC public schools. They prepared this analysis for our Network.

THE *CRISP* DATA MINING PROCESS

01 BUSINESS UNDERSTANDING

Understand the **problem** to be solved.

02 DATA UNDERSTANDING

Understand the **strengths and limitations** of the data.

03 DATA PREPARATION

Manipulate and convert the data into forms that yield better results.

04 MODELING

Apply **data mining techniques** to capture regularities in the data.

05 EVALUATION

Assess the data mining results rigorously (with stakeholders) and **gain confidence** that they are valid and reliable before moving on.

06 DEPLOYMENT

Put the results into **real use** in order to realize some return on investment.

SOCIAL-EMOTIONAL LEARNING (SEL)

SEL consists of mindsets, behaviors, and skills a person needs to succeed

MINDSETS	BEHAVIORS	SKILLS
Growth Mindset	Self-Advocacy	Interpersonal Skills
Academic Self-Efficacy	Academic Behaviors	
Belonging	Problem-Solving	

We measure SEL in students using a 53-item survey that members administer twice a year

PRE-SURVEY (baseline)

Administered at the start of the school year

POST-SURVEY (change)

Administered at the end of the school year

~5000 baseline surveys and **~5000 post-surveys** from students in 5th grade to college at 210 sites were included in this bright spot analysis

BUSINESS UNDERSTANDING

What practices affect **growth** in student social-emotional learning?

What can we **generalize** about the practices and characteristics of sites that **grew more than their peers** in SEL?
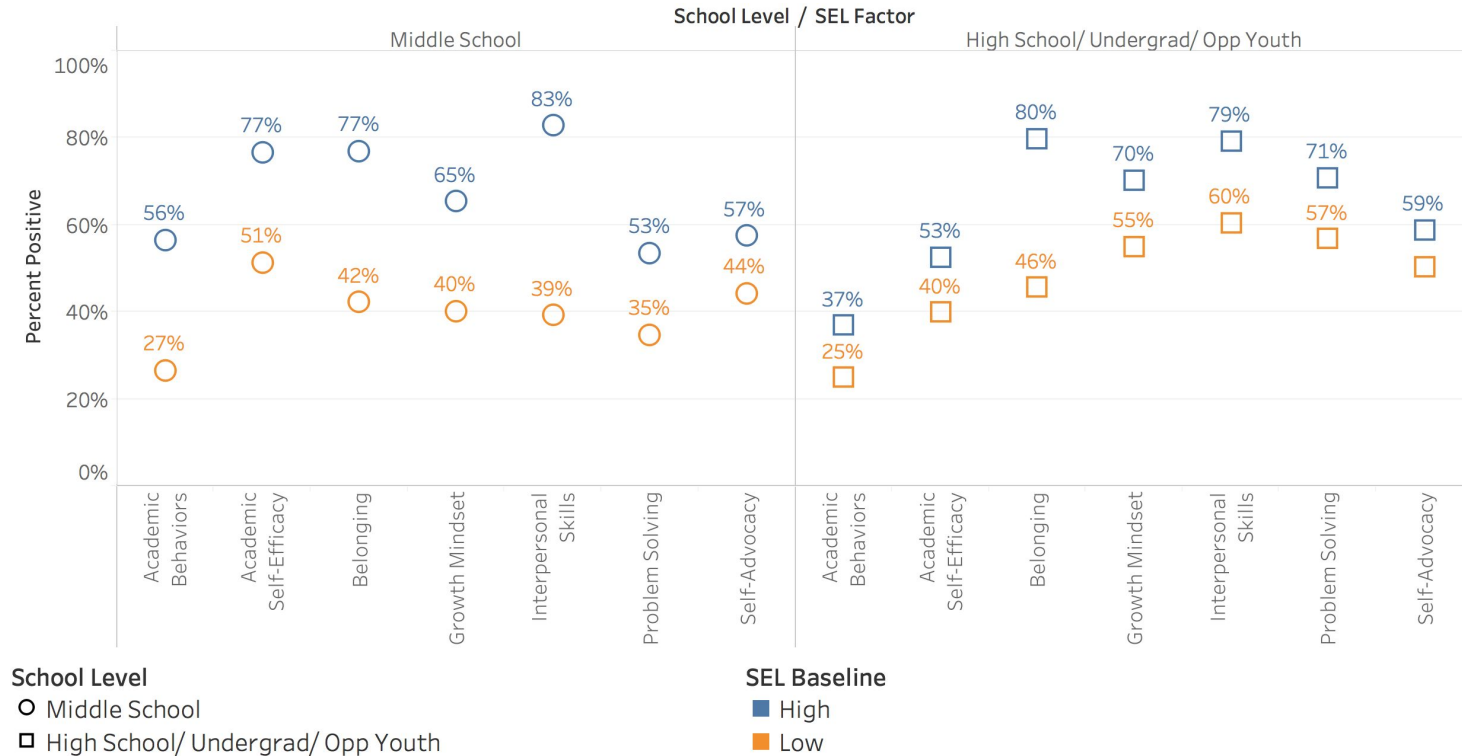
MODELING

We used a K-means procedure to cluster our data.

1. Started with a **hypothesis:** a student's post-survey SEL score has something to do with their baseline SEL score -- where they started -- and their age group.

2. **Reshaped** the data and chose 5 random points as initial centers.

3. **Created clusters** by identifying the data points closest to the initial centers.

4. **Recalculated** the centers by finding the *centroids* of the clusters created in step 3.

5. Ran the clusters through several **statistical tests** to determine their quality.

6. **Repeated steps** until we landed on the correct number and quality of clusters.

MODELING

We landed on 4 clusters:

School Level / SEL Factor

MODELING

We applied a multilevel statistical model to determine which sites deviated farther from their cluster estimate in the post-survey.

Why? Our data has hierarchy:

- Student → Site → Organization → NYC

- The SEL scores of students within a site should be *more similar* to their peers in the same site and *more different* from students in another site. This is good for comparison.

- However, *all* SSN members operate within NYC, which gives them a lot in common.

DEPLOYMENT

We found **10 sites** that deviated significantly from their cluster. Not every SEL cluster had a bright spot site, and some sites repeated.

Then, we interviewed people at those sites using an interview protocol to uncover **practices**.

The people then shared their practices on a **panel** at a Network-wide event.

Each circle represents an organization with a bright spot site.