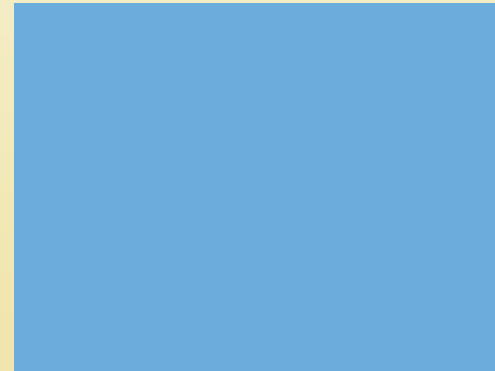


# Analítica Avanzada y su impacto en educación

0101  
01010  
010  
01010

NEAL  
ANALYTICS

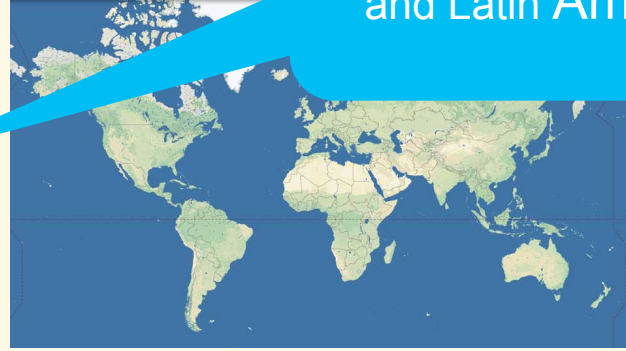


# Jorge P-LUNA intro



at&t

Director IT Europe  
and Latin America



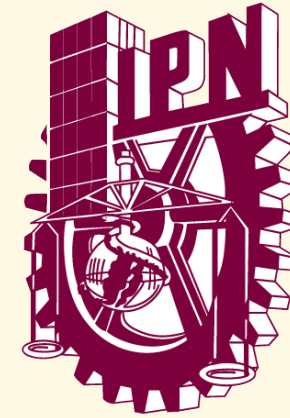
**NEXTEL**®

nextel®

VP / Chief  
Information Officer  
INTERNATIONAL

BILL & MELINDA  
GATES foundation

Senior Technology  
Officer  
Global  
Development



Computer Science



MBA  
International  
Business

# Pasión por Ciencia De Los Datos



**Agricultura**: Podemos predecir cuando la tierra dejara de producir por erosión?



**Servicios Financieros**: Podemos predecir si una persona que no tiene una “reputación” pagara un préstamo de banco para un negocio nuevo?



**Educación**: Podemos predecir si una persona de pocos recursos puede terminar la Universidad? Como ayudarlo ?

# Neal Analytics: Microsoft Partner



## Our Mission

Drive customer value with Analytics on the Microsoft platform, and to drive IT & business partnership in our clients. We utilize a host of Microsoft technology to drive this value.

## Our Company

We are Seattle-based company with 30 data engineers and scientists that have helped dozens of customers improve their businesses. We were founded in 2011.

## Manu., Retail, Edu.

Our objective is to make analytics accessible to organizations of all sizes across our verticals. Our team specializes in creation of analytical practices to help institutions grow and scale.

## Partnerships

We are a Microsoft partner that develops solutions on using Azure ML, Azure marketplace, HDInsight, Stream Analytics, Azure Data Factory, SQL Server and Event Hubs.

## Our Focus

Demand forecasting, decision modelling, resource forecasting, predictive maintenance, systems integration and creating more profitable customers.

## OUR PROCESS

Prove ROI

Broadly Engage Leaders

Production

Solution Sales Director: David Brown, 425-283-6842,  
davidb@nealanalytics.com  
<http://www.nealanalytics.com>

**Microsoft Partner**  
Gold Data Analytics  
Gold Cloud Platform

Microsoft Partner of the Year  
2015 Finalist  
Big Data and Analytics



# Objetivos de un proyecto AA

05 Reducir deserción y abandono



04 Cierre de la brecha de digital y educativa en la educación



03 Fortalecer la educación pública.















02 Mejorar la calidad de aprendizaje



01 Revalorizar la Carrera Docente.



# Posibles Escenarios

<p>Que estudiantes están en riesgo de dejar la escuela</p> 	<p>I Quiero saber quienes son mis mejores profesores!</p> 	<p>Cual es la estrategia correcta para una intervención de un niño en riesgo?</p> 	<p>Quiero optimizar mis recursos en la educación!</p> 
<p>Tenemos toda esta data pero no hacemos nada con ella!</p> 	<p>Quiero una mejor ganancia en mis programas educativos!</p> 	<p>Quiero un Tablero de control que me indique por profesor y niño/niña su rendimiento!</p> 	<p>Quiero medir por profesor el riesgo de que niños no aprendan ____!</p> 
<p>Que estudiantes están muy abajo o muy arriba en su desempeño y por que?</p> 	<p>Quiero una mejor manera de crear innovación escolar!</p> 	<p>Quiero mejorar mi Curriculum</p> 	<p>Quiero saber que estudiantes son genios.</p> 

# Aplicación de Modelos para cumplir objetivos



Mejorar Calidad



Revalorización de la carrera docente



Fortalecer La Educación Pública



Reducir Deserción Y abandono

Adopción, Implementación, Manejo del cambio, adaptación de los sistemas de información, entrenamiento

5

Process Engineering

Diseñar nuevos programas de educación, entrenar a los profesores, mejorar salaries, etc.

4

Predictive Models

Modelo para predecir niños que saldrán escuela

3

Business Scenarios

Estudiantes en riesgo, quien es mi mejor profesor?, cual es el mejor programa académico para un niño?, etc.

2

Business Intelligence

Calificaciones, tareas, exámenes, etc.

1

Source Data

Estudiantes, Profesores, GEOData, Gobierno, Social Media, etc.



# Oportunidades

## Ciclo de vida del análisis de datos

- Tomar decisiones basadas en los resultados
- Definir estándares de servicio
- Metas para alcanzar estándares
- Planificar y asignar recursos

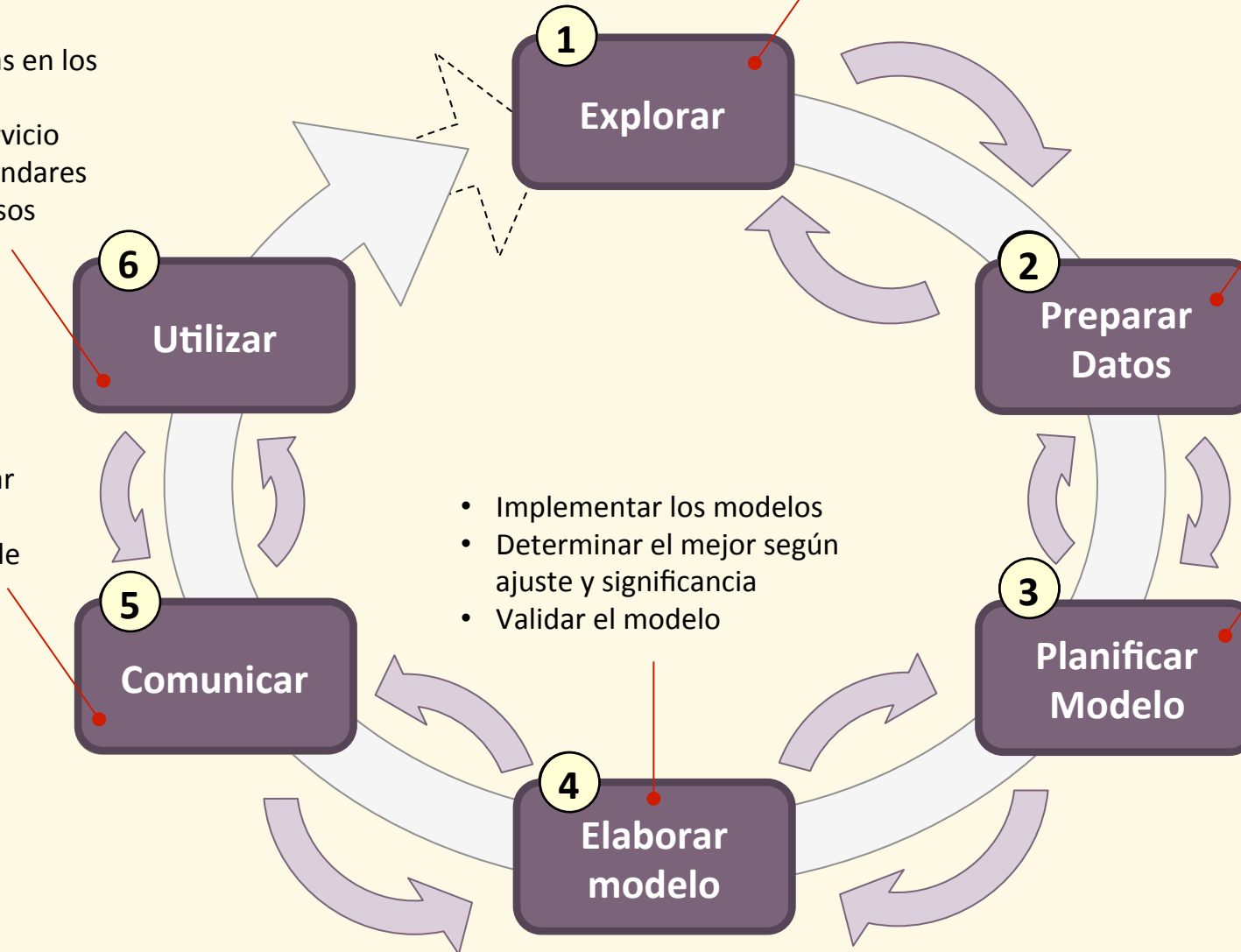
- Interpretar resultados
- Generar visualizaciones adecuadas para comunicar resultados
- Hacer recomendaciones de mejora

- Explorar los datos disponibles
- Formular la pregunta
- Formular hipótesis

- Determinar datos necesarios
- Recopilar datos
- Limpiar datos
- Analizar consistencia de datos

- Determinar qué variables explican y cuáles predecir
- Seleccionar posibles modelos y algoritmos a utilizar
- Definir métricas de desempeño

- Implementar los modelos
- Determinar el mejor según ajuste y significancia
- Validar el modelo





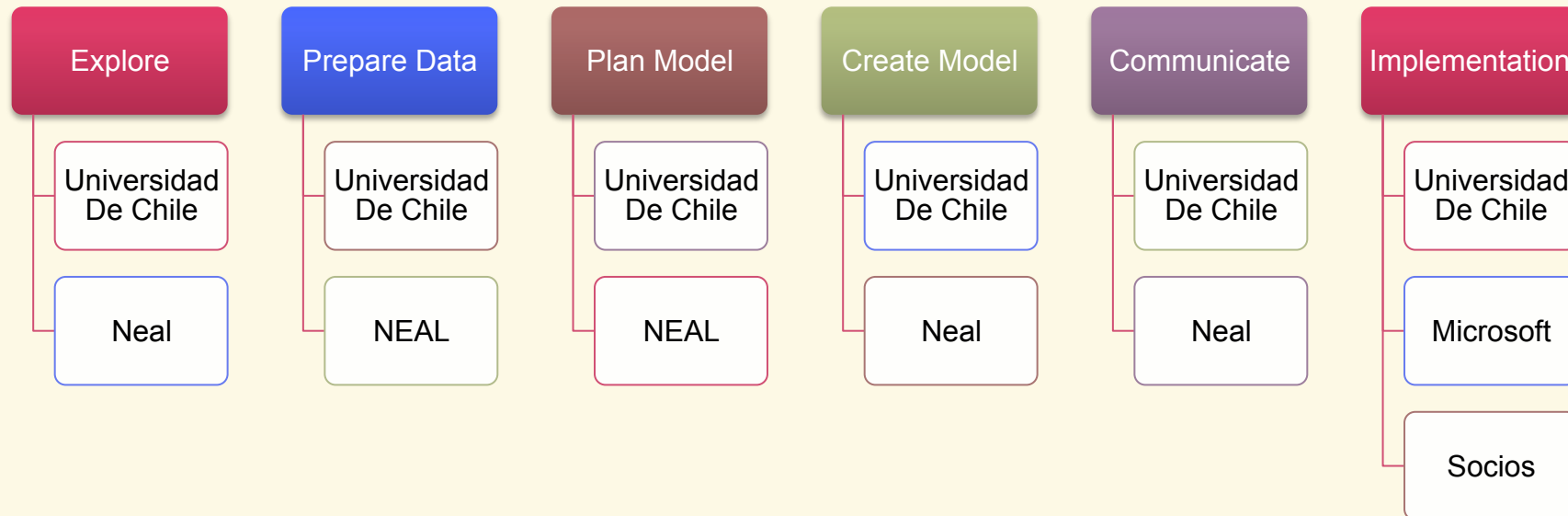
# *Casos en Educación*

## *Predicción de la deserción y abandono escolar*



# Alianza Universidad De Chile, Microsoft, Neal

1. Trabajamos juntos en la transferencia de conocimientos de herramientas Microsoft a la Universidad.
2. En equipo elaboramos un plan detallado.
3. Creamos en detalle los requerimientos y escenarios.



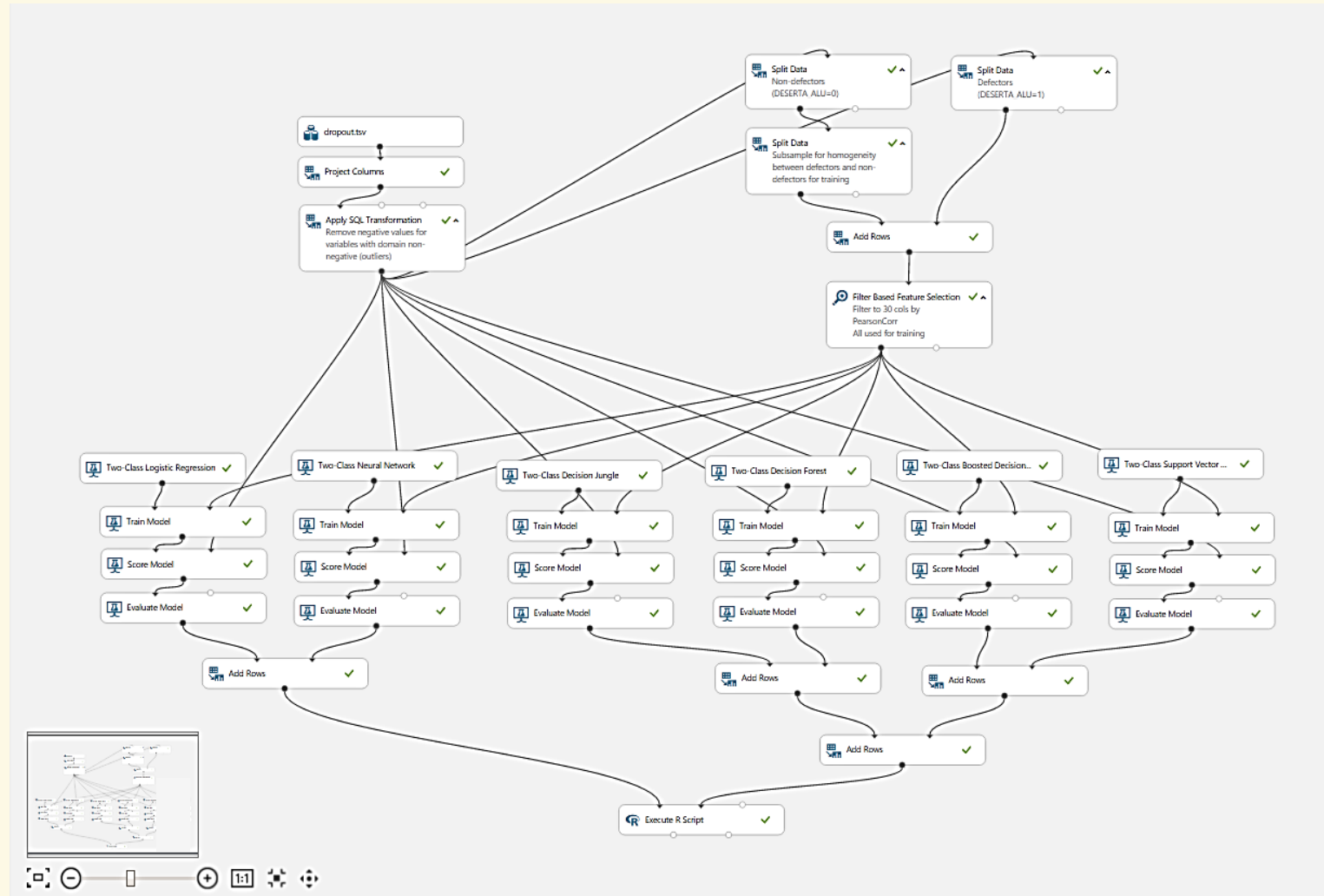
# Modelos explicativos y predictivos

## Modelos predictivos: Azure ML

- Predicción de la **deserción y abandono** escolar:
  - Se usaron **127** atributos de los estudiantes, establecimientos y las manzanas donde viven o se ubican para alimentar el algoritmo de aprendizaje automático
  - Se utilizó Azure ML para identificar los atributos más significativos reduciéndolos a **31**
  - Aparecen significativas variables como la vulnerabilidad del colegio, la convivencia, participación, autoestima y motivación como factores que permiten **explicar** y **predecir** la deserción y abandono escolar

# Modelos explicativos y predictivos

## Modelos predictivos: Azure ML



# Modelos explicativos y predictivos

## Ajuste del modelo

Microsoft Azure  
Machine Learning

Dropout Model (defectors) ▶ Execute R Script ▶ Result Dataset

rows 6  
columns 8

Algorithm	Accuracy	Precision	Recall	F-Score	AUC	Average Log Loss	Training Log Loss
Logistic Regression	0.925667	0.266003	0.860471	0.406379	0.951995	0.235705	-76.901555
Neural Network	0.029705	0.029573	1	0.057447	0.860224	9.844628	-7288.585055
Decision Jungle	0.926418	0.274812	0.908246	0.421952	0.972655	0.204718	-53.644733
<b>Decision Forest</b>	<b>0.937317</b>	<b>0.318159</b>	<b>0.979712</b>	<b>0.480331</b>	<b>0.985494</b>	<b>0.560043</b>	<b>-320.32344</b>
Boosted Decision Tree	0.927707	0.282431	0.937827	0.434124	0.979829	0.206087	-54.672232
SVM	0.847723	0.068369	0.328665	0.113192	0.596334	0.552368	-314.563072

### ALGORITHM BEST FIT

#### Two-Class Decision Forest

True Positive  
**7581**

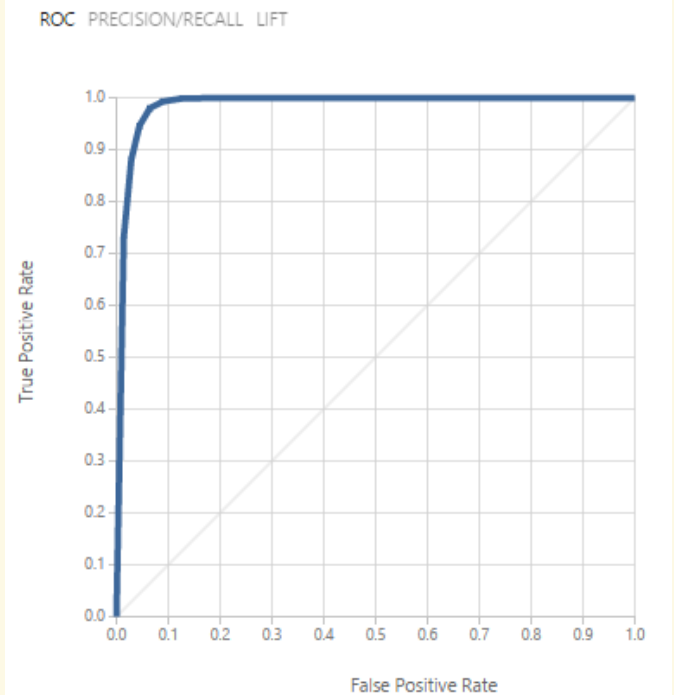
False Negative  
**59**

False Positive  
**22370**

True Negative  
**228368**

**Confusion matrix**

	1	0
1	99,23%	0,77%
0	8,92%	91,08%



# Casos en Educación

## “Restart Academy of Missouri”

# Scenario: “Restart Academy of Missouri”

- Restart Academy (RA) is a “second chance” school in Missouri, for 5,625 students who have failed in their original schools and are attempting to restart their educations.
- Students at all public school grade levels participate in RA. Classes at RA are taught on line, except in St. Louis and Kansas City MO where there is a high enough concentration of students, for conventional class room teaching.

School Type	Online	In-Class
Grade school (K-5)	North Elementary South Elementary	St. Louis Elementary Kansas City Elementary
Middle school (6-8)	North Middle South Middle	St. Louis Middle Kansas City Middle
High School	North High South High	St. Louis High Kansas City High

12 schools total  
5,625 students

- Task is to predict who is likely to drop out from student data, and then.
- What drop-out prevention intervention, is most likely to succeed for each student.



# Scenario Users

## Teachers



- Managing individual students
- Interest focused on
  - “How are my students doing?”
  - “Which students are at risk for dropping out?”

## Principals



- Managing schools and teachers within schools
- Interest focused on
  - “How is my school doing?”
  - “How are my teachers doing?”
  - “How many at risk students do we have?” “Which teacher has the most potential drop outs?”

## Superintendent/ CIO



- Managing a set of schools
- Interest focused on
  - “How is my school system doing?”
  - “How is online vs. in-class teaching going?”
  - “What will it cost to handle drop outs?”

# The Problem



- Average high school dropouts
  - earn \$20,241 per year, according to the [U.S. Census Bureau](#).
  - \$10,386 less than high school graduates, and
  - \$36,424 less than college graduates.

Source: <http://content.time.com/time/covers/0,16641,20060417,00.html>

# Four Solutions

## Four Core Dropout Prevention Strategies

1. After School Opportunities
2. Alternative Schooling
3. Mentoring/Tutoring
4. Service-Learning



Programs of Study as State Mandate: A Longitudinal Study of the Personal Pathways to Success Initiative

[Read the Report](#)

NATIONAL DROPOUT PREVENTION CENTER / NETWORK  
SUPPORTING LEARNERS AND TRANSFORMING LIVES

**NRC** CTE  
National Research Center for Career and Technical Education

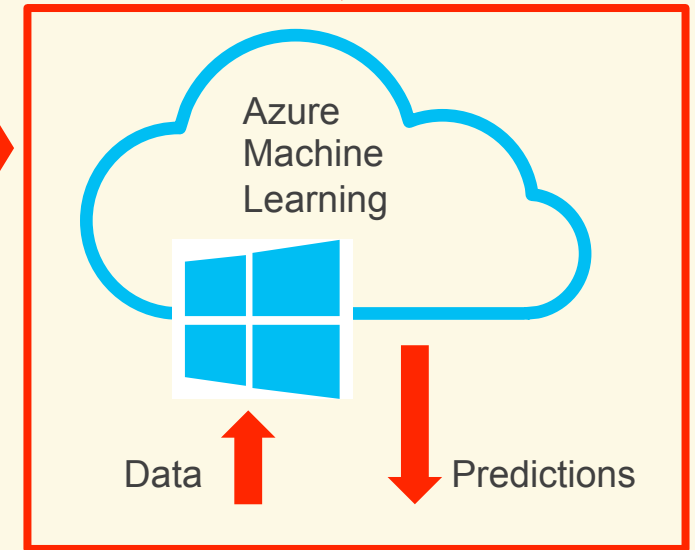
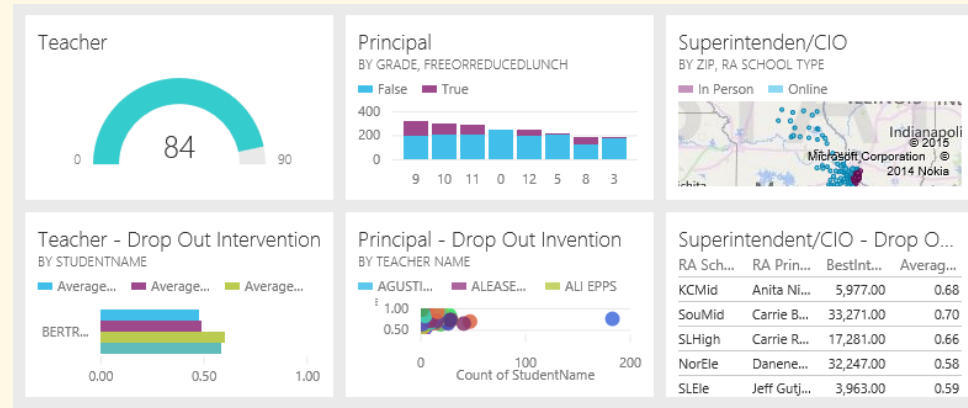
<http://www.dropoutprevention.org/> At Clemson University

Easy Connection To Azure

### Source Data

- Historical performance
- Student demographic and
- Diagnostic
- Disciplinary
- Extra-Curricular

Plug Existing Data From Excel ... Into Machine Learning



Power BI Visualization



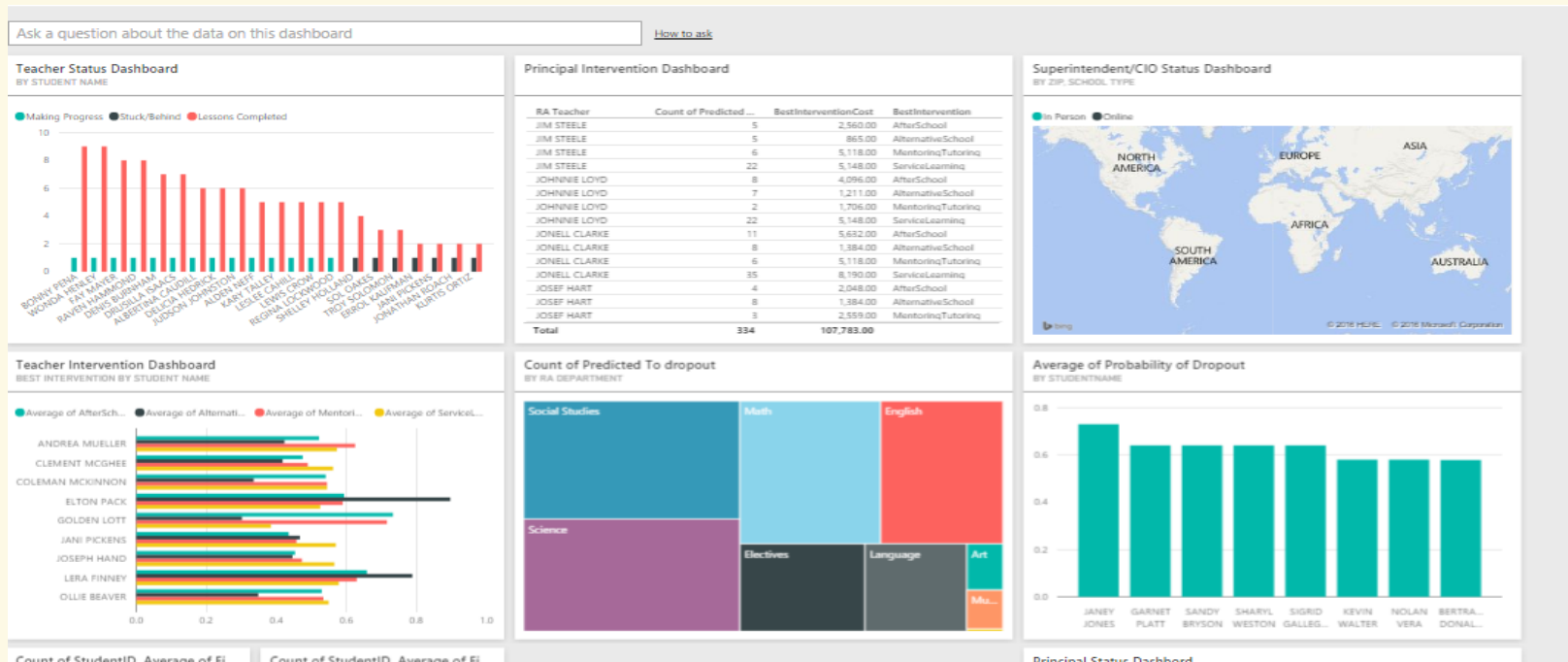
Excel

Existing Worksheets

# Ejemplo de análisis por profesor

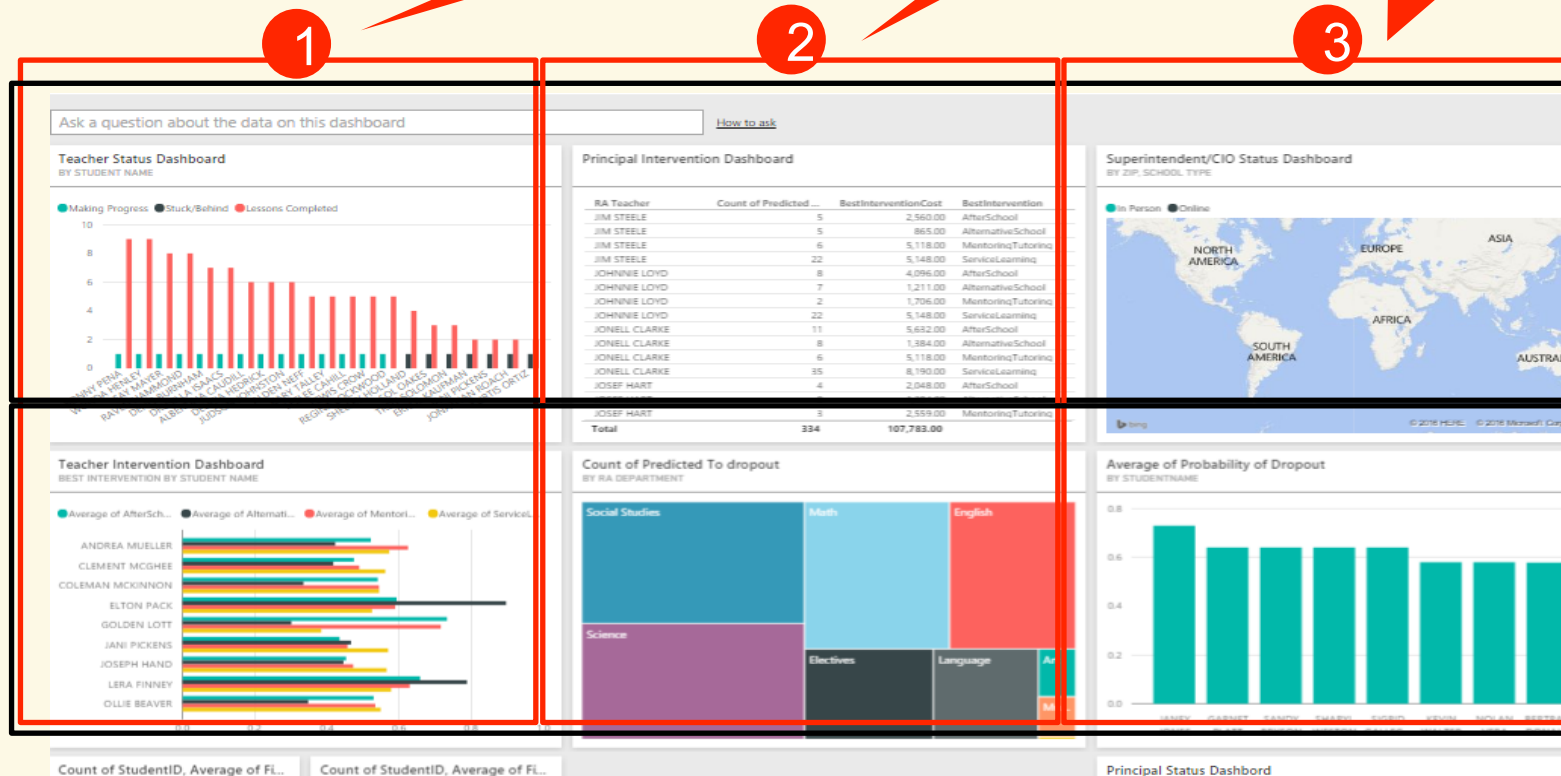
## Maestro

- A. Como esta mi clase?
- B. En que área están mis alumnos atrasados?
- C. Que estudiantes están en riesgo de salir?
- D. Cual es la mejor estrategia para evitar salga de la clase?



# Crear Tableros de información para intervención usando ML

Profesores, Subdirectores, Directores de área

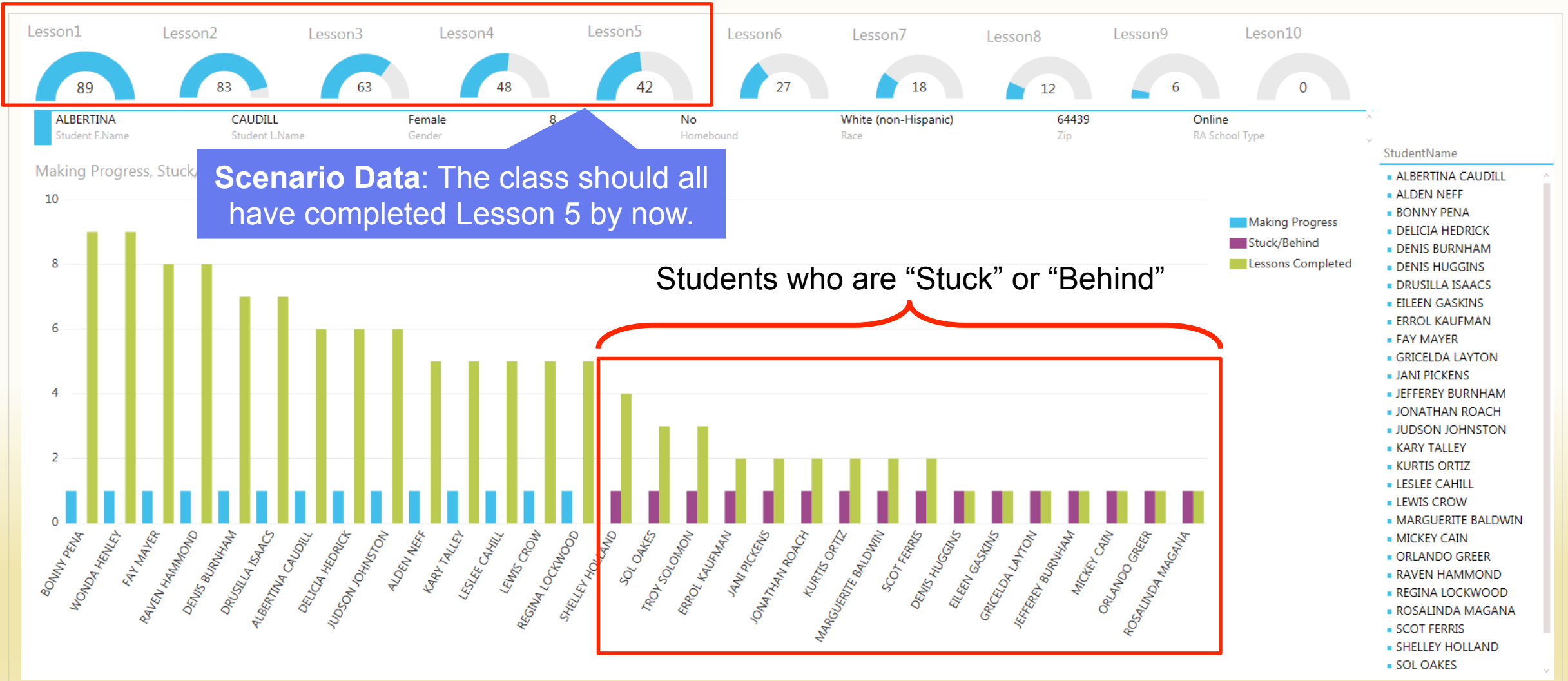


Como van mis alumnos?

Que intervención puedo dar?

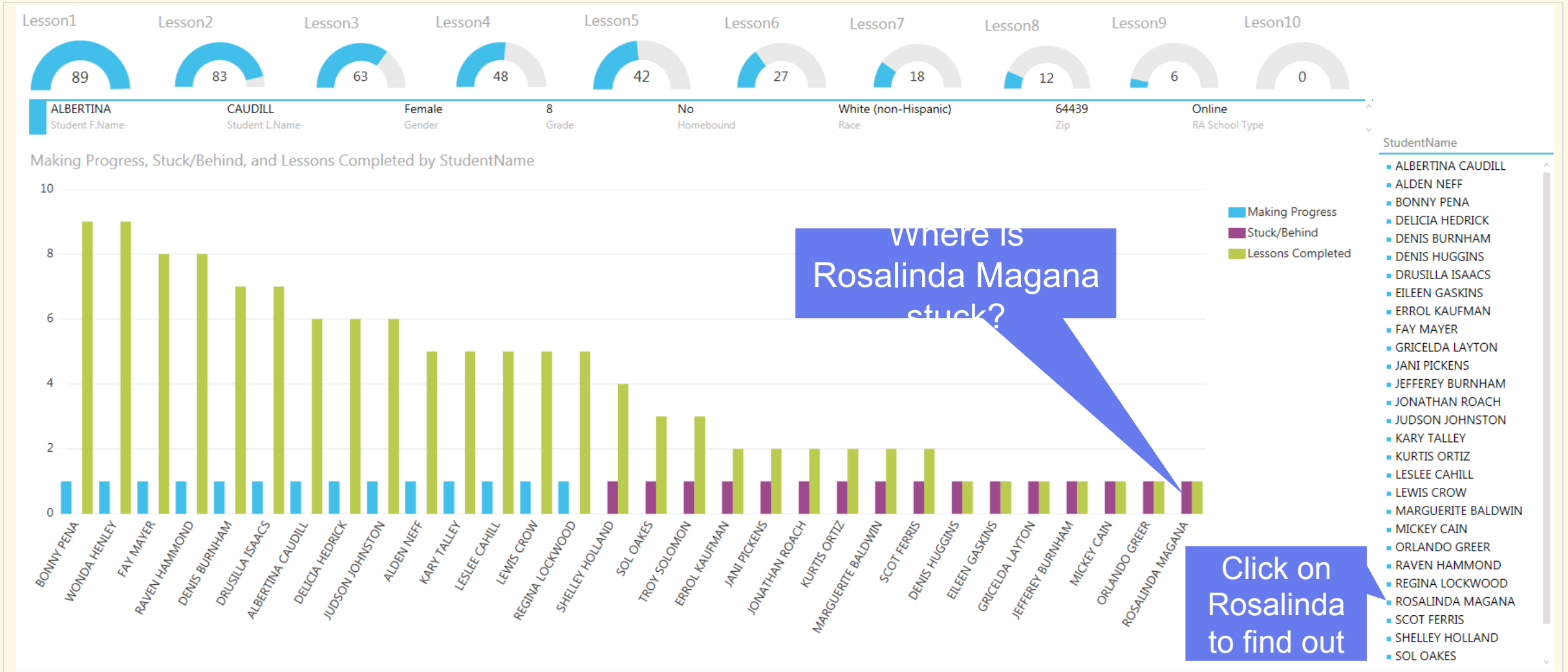


# Tableros por cada profesor y sus niños:





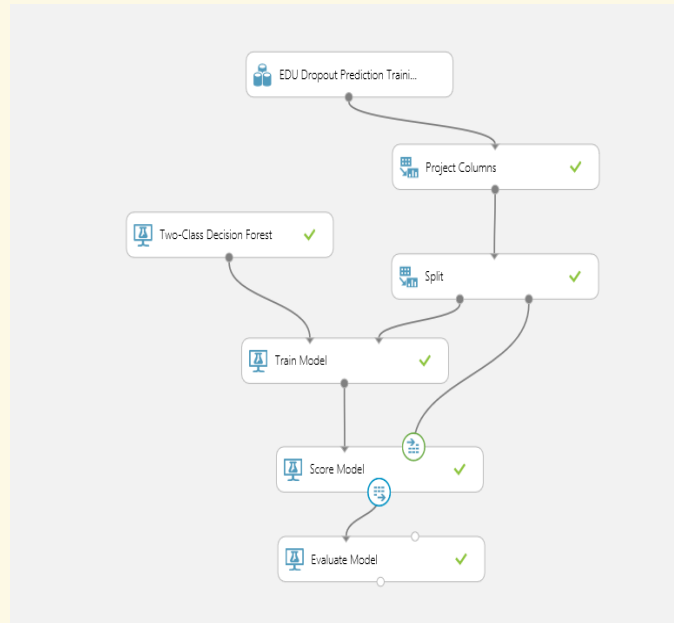
# Tablero por profesor: Como esta mi clase?



# Modelo Predictivo usando Azure ML

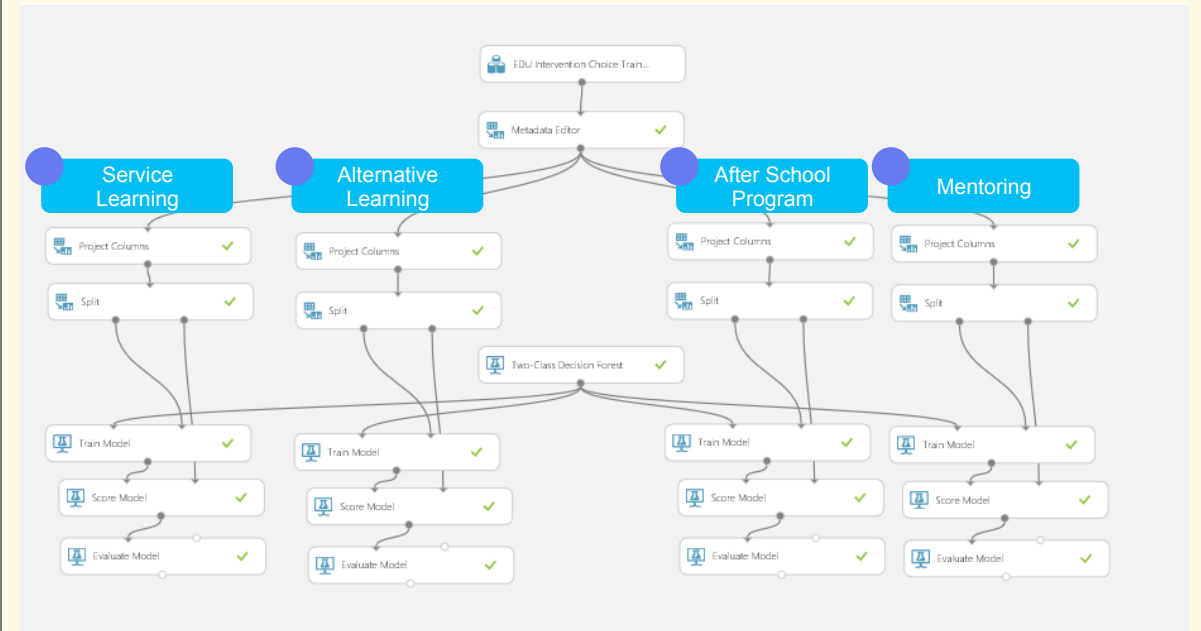
# Predictive Models Overview

## Drop Out Prediction Model



- Single model that predicts drop out on individual student basis
- Model is web-service ready. – Indicated by green and blue circles
- Outputs of prediction include binary 1/0 prediction whether student will drop out and relative certainty of prediction

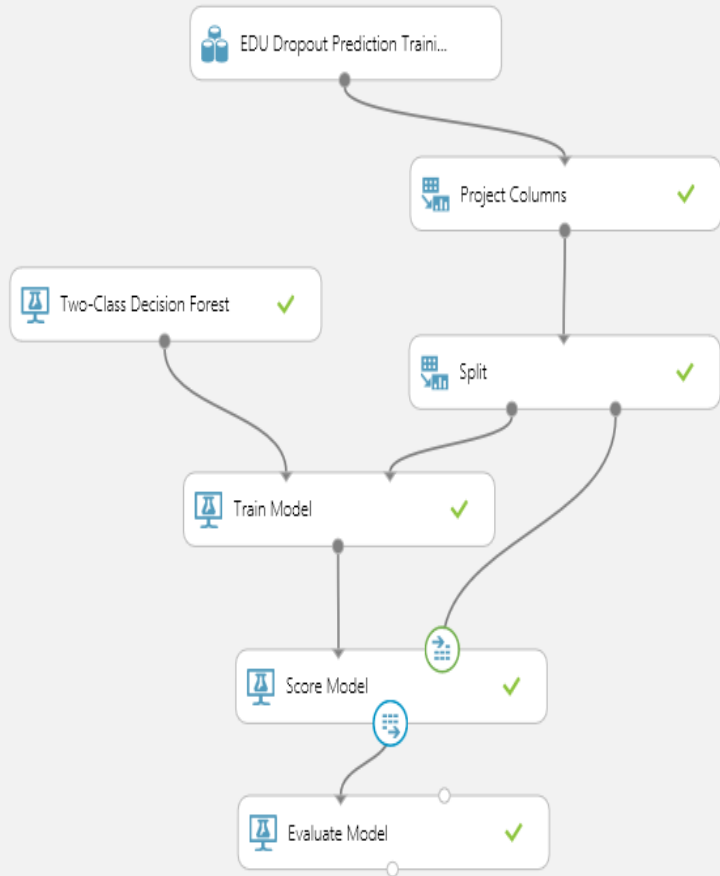
## Intervention Effectiveness Model



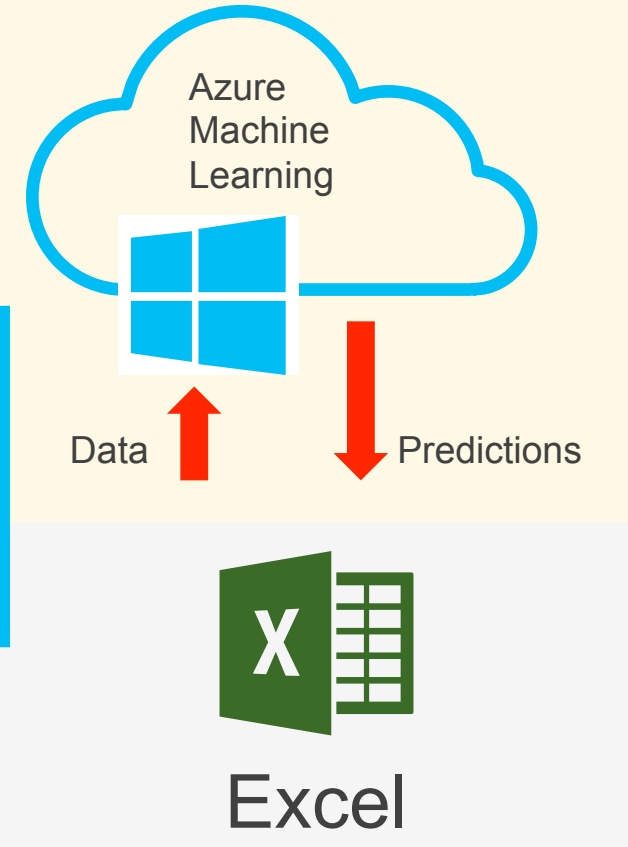
- Model composed of four prediction sub-models. Each sub-model creates a prediction for an intervention method
- Output of model includes binary 1/0 prediction of success and relative certainty 0-1.0 rating for each intervention
- Intervention methods covered:
  - Service Learning
  - Alternative Learning
  - After School Programs
  - Mentoring

# Drop Out Prediction Model

Azure Machine Learning Model

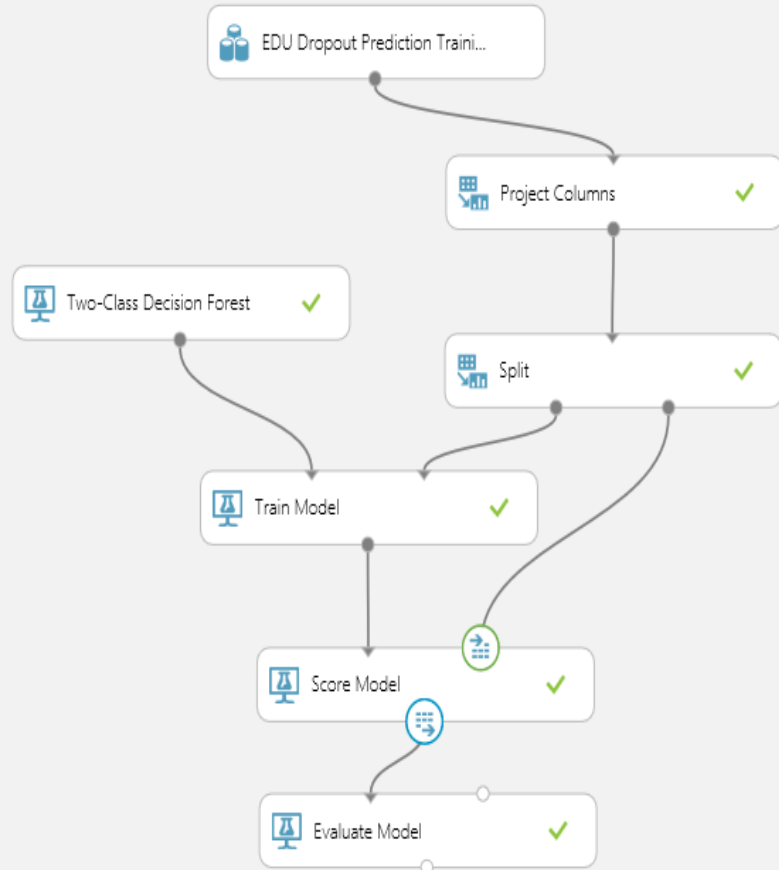


This drag and drop canvas is how Azure Machine Learning Builds Models. Anyone in the PTO can do this! Teachers, Administrators, Parents,



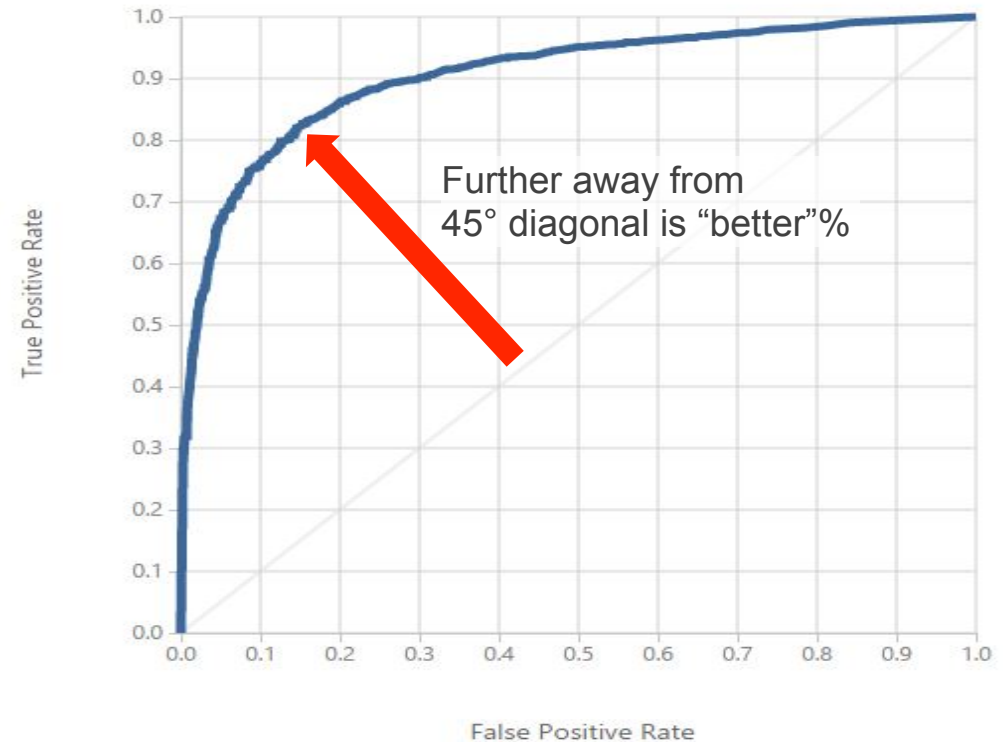
# How Good Is Our Drop Out Prediction Model?

## Existing Worksheets



## EDU Demo Dropout predictor > Evaluate Model > Evaluation results

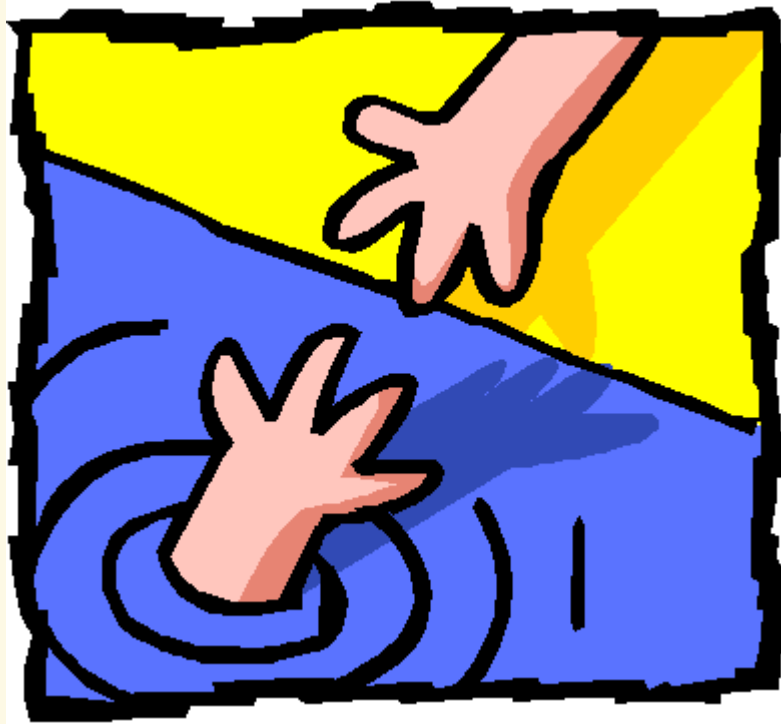
ROC PRECISION/RECALL LIFT



*Azure Machine Learning Model is 80% Accurate predicting who will drop out from class data*

# Great, Now We Know Who Will Drop Out

- But, ... can machine learning do more?
- For Example:  
Can we predict which of the four dropout intervention strategies will be best for each individual student?



# Interventions View

Based on Historical Data ...  
Predictions can be made for each  
of the 4 Clemson strategies, for  
each student

## Basic Core Dropout Prevention Strategies

1. After School Opportunities
2. Alternative Schooling
3. Mentoring/Tutoring
4. Service-Learning



National Dropout Prevention Center/Network



Programs  
of Study as  
State  
Mandate:  
A Longitudinal  
Study of the  
Personal Pathways  
to Success Initiative

[Read the Report](#)

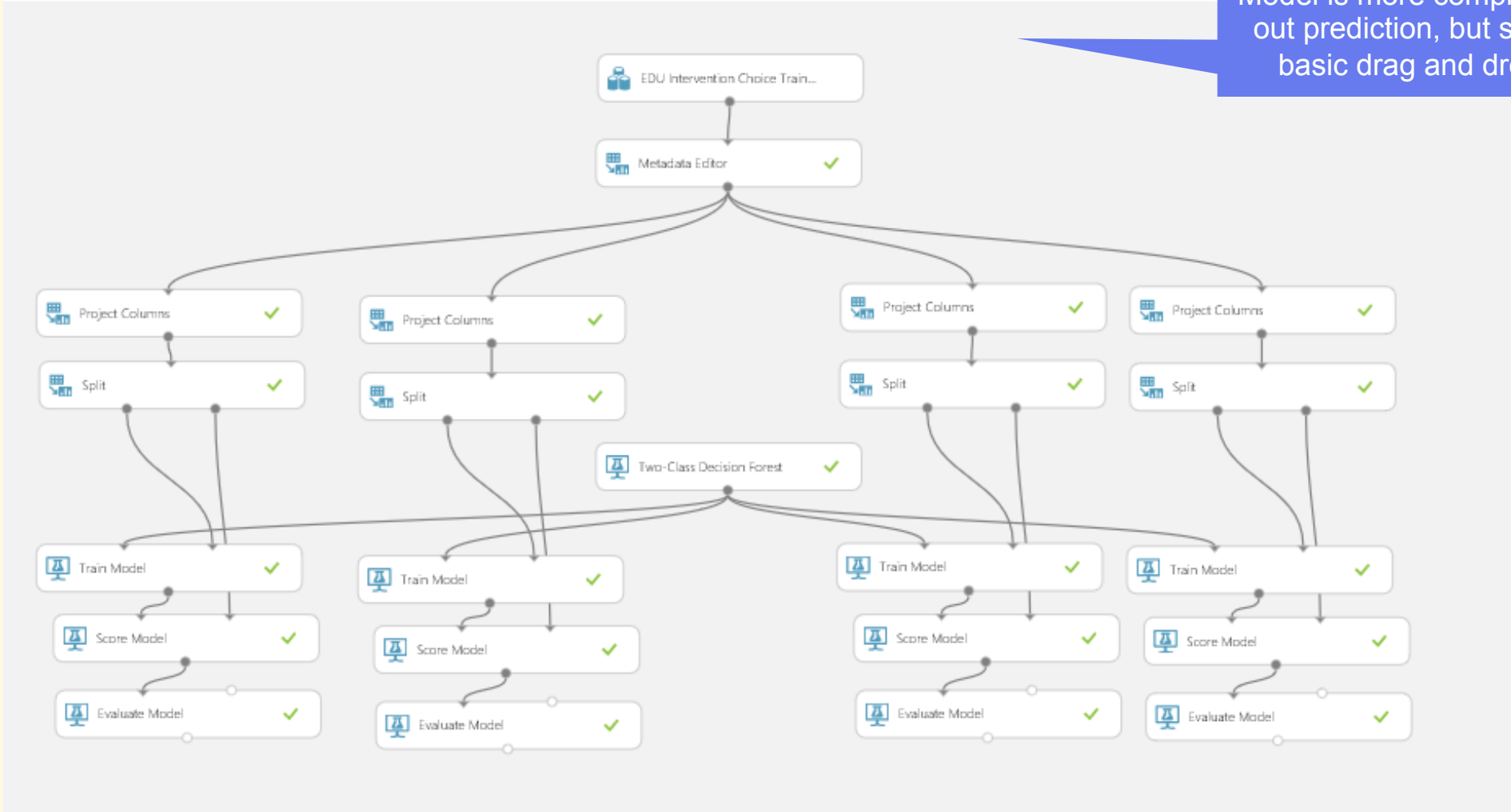
**NRC** CTE  
National Research  
Center for Career and  
Technical Education



# Yes, ... By Building An Intervention Prediction Model

EDU - Intervention Method Success Prediction

Model is more complex than drop-out prediction, but still the same basic drag and drop to-build



# Teacher Intervention Dashboard

- Basic Core Dropout Prevention Strategies
1. After School Opportunities
  2. Alternative Schooling
  3. Mentoring/Tutoring
  4. Service-Learning



2

**Predicted to Dropout**

0  
 1

**AILENE CROWLEY**

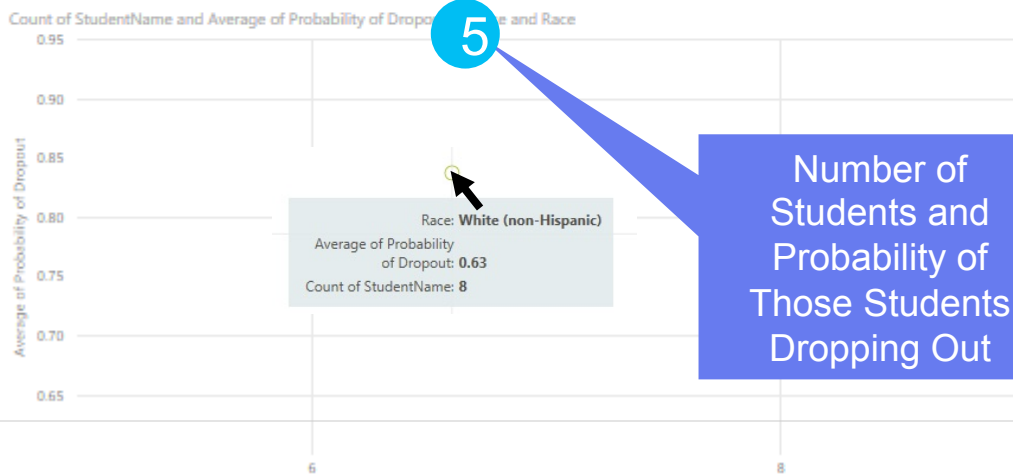
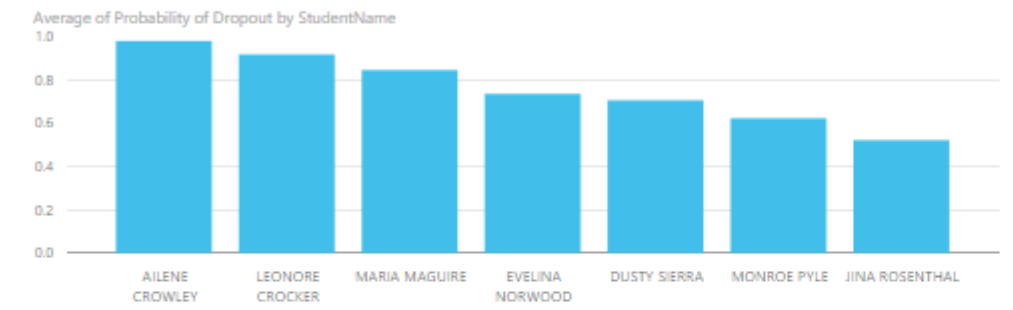
0.67 Average of AfterScho...  
0.89 Average of Alternativ...  
0.70 Average of Mentoring...  
0.59 Average of ServiceLearnin...

7 Count of StudentName | 1 Count of Teacher Name | 5 Count of RA Principal

Average of AfterSchoolProb, Average of AlternativeSchoolProb, Average of MentoringTutoringProb and Average of ServiceLearningProb by Stud...

- Teacher Name**
- AGUSTIN LOOMIS
  - ALEASE DUDLEY
  - AMI LEPPS
  - ANTON STERN
  - ANTHONY HASKINS
  - ANNA SPARKS
  - ANTONIO CARVER
  - ARNOLD CHACON
  - BENEDICT NORTH
  - BERNARD SPEARS
  - BOB ECHOLS
  - BODDY BERNIER
  - BRADFORD ELLIOT
  - BRUNELINE SCHMITT
  - BRUNO HAMM
- Select Teacher

3

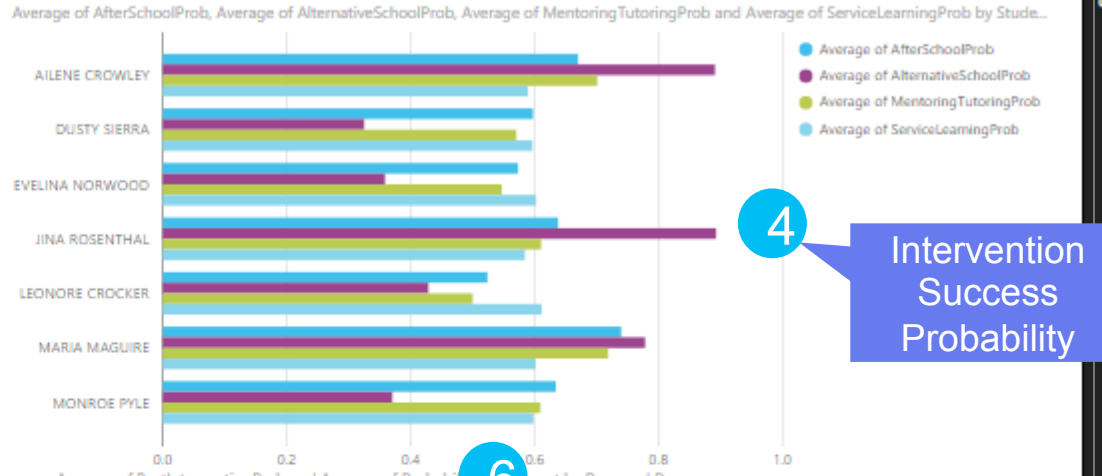


5

Number of Students and Probability of Those Students Dropping Out

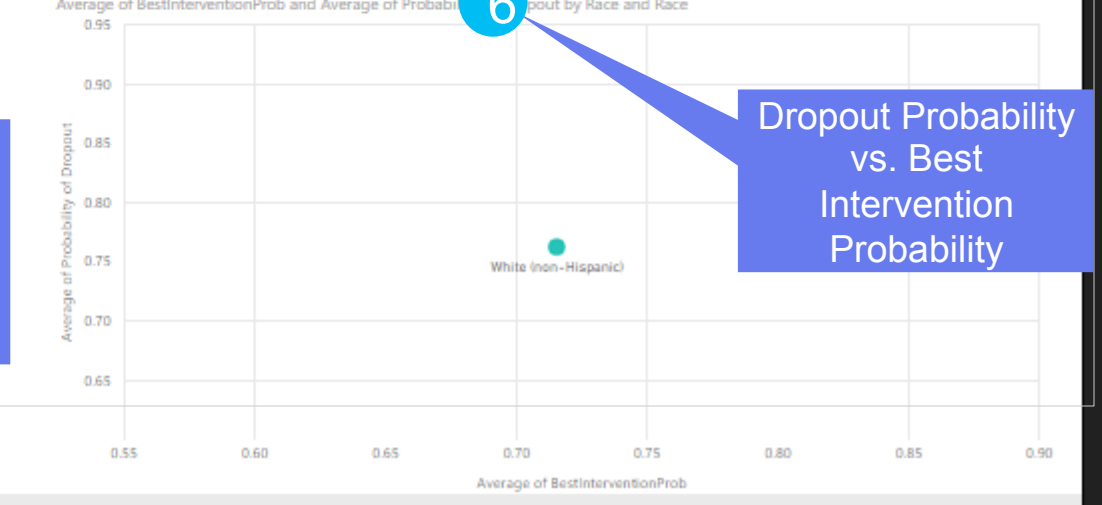
4

Intervention Success Probability

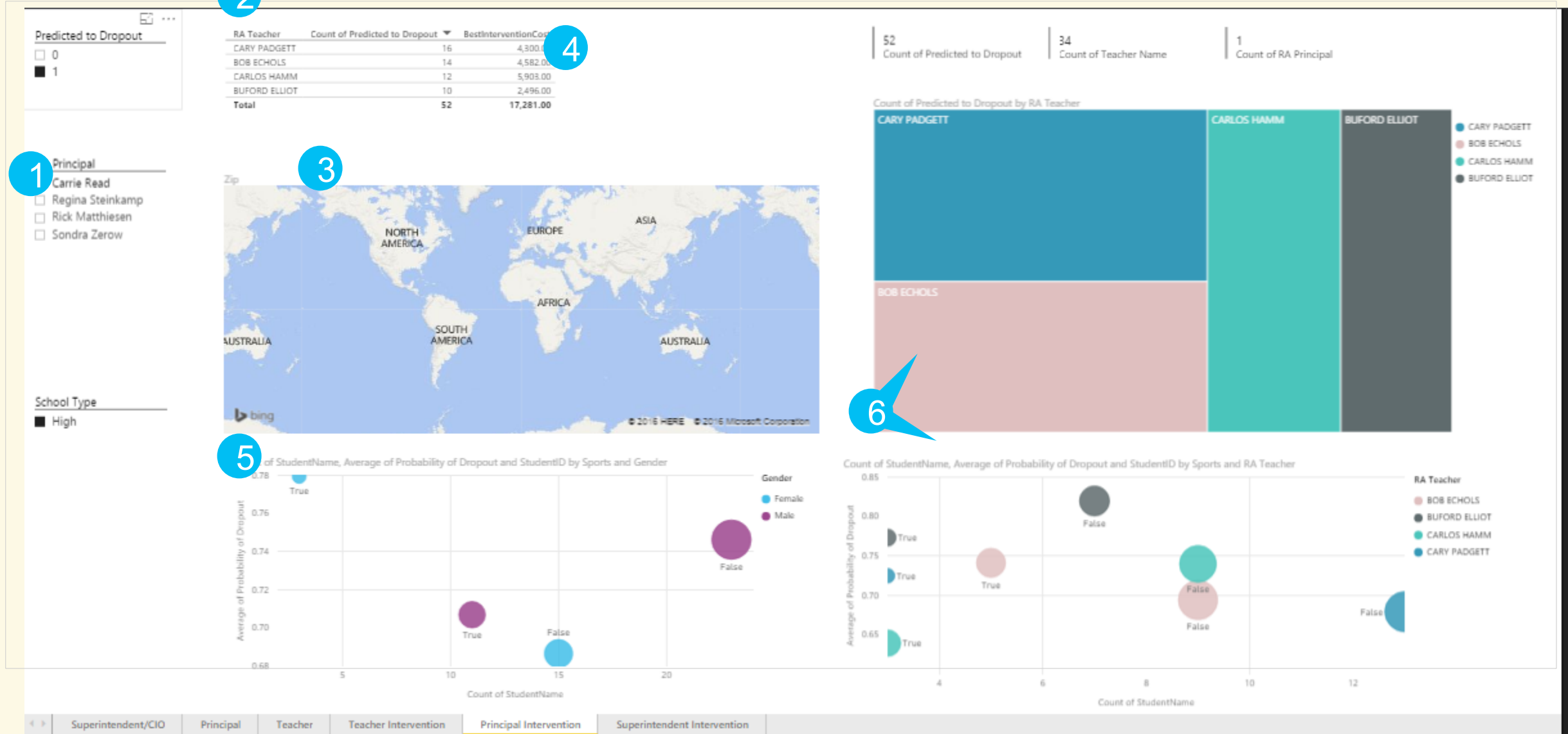


6

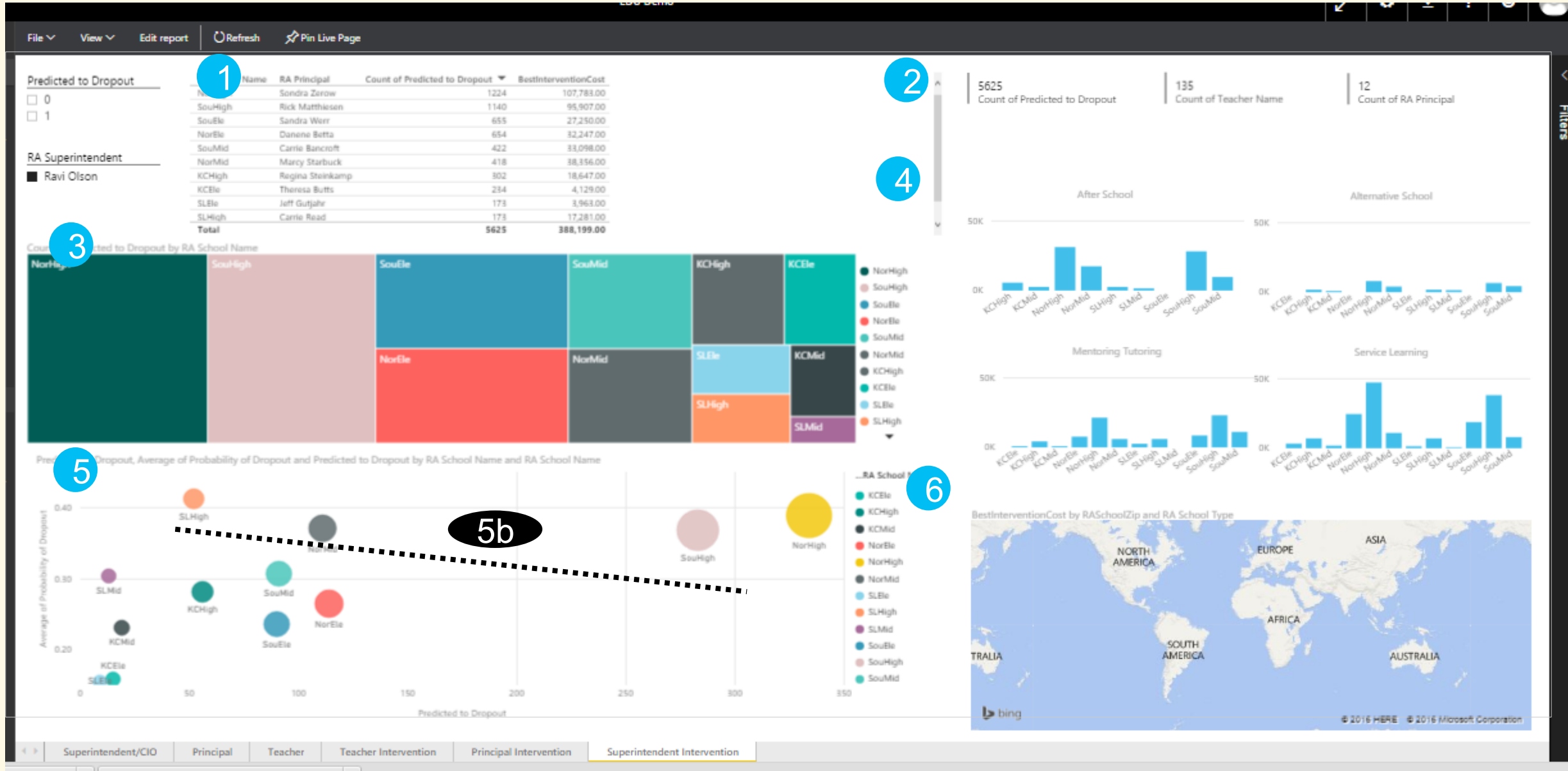
Dropout Probability vs. Best Intervention Probability



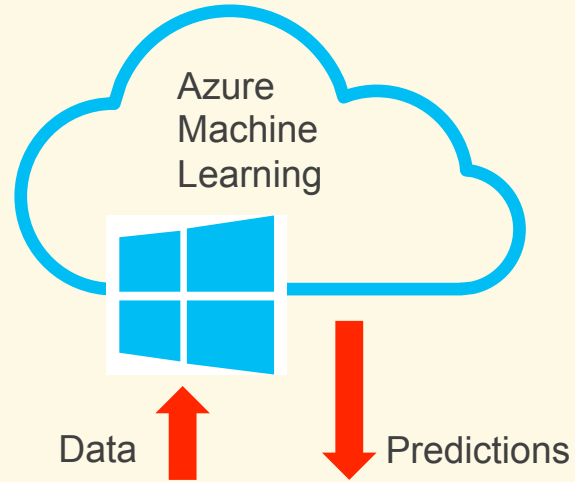
# Principal Intervention Dashboard



# Superintendent Intervention Dashboard



# The Journey



## Student Info

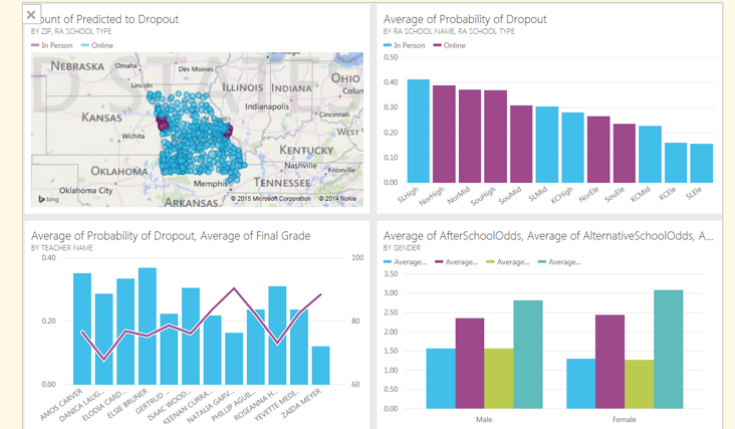
- 10 Variables
- Starting life in teacher grade books

## Azure Machine Learning

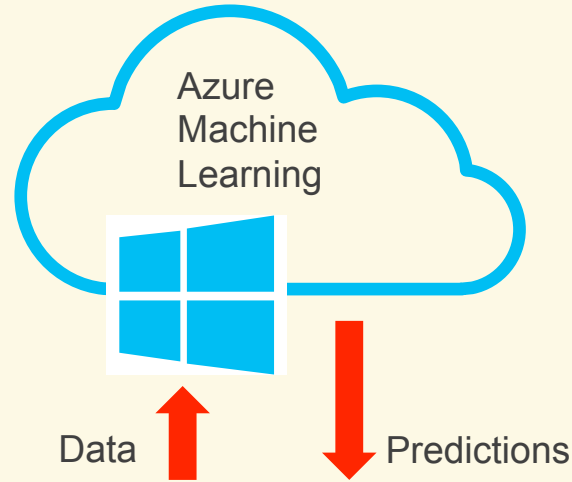
- Data to cloud ... via Excel
- Bring back predictions

## PowerBI

- Self-Serve
- Data Visuals



# The Journey In 3 Steps



## Student Info

- 10 Variables
- Starting life in teacher grade books

## Azure Machine Learning

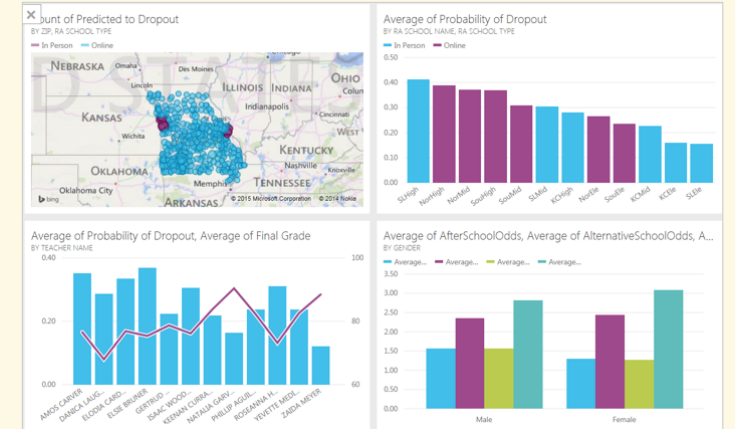
- Send to cloud via Excel
- Bring back predictions

## PowerBI

- Self-Serve
- Data Visuals

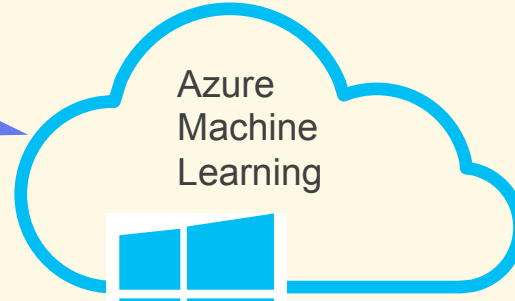


Step1: Start with tool you know well, and data you already have



# The Journey: Step 1

Step2: Leverage the most powerful cloud on the planet



Data



Predictions



## Student Info

- 10 Variables
- Starting life in teacher grade books

## Azure Machine Learning

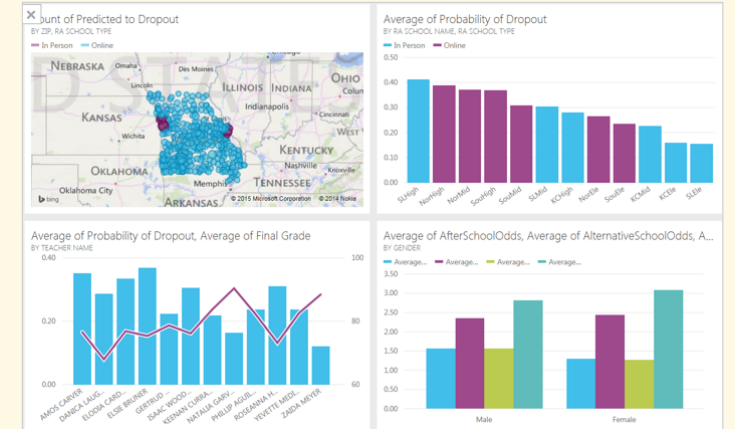
- Send to cloud via Excel
- Bring back predictions

## PowerBI

- Self-Serve
- Data Visuals



Step1: Start with tool you know well, and data you already have

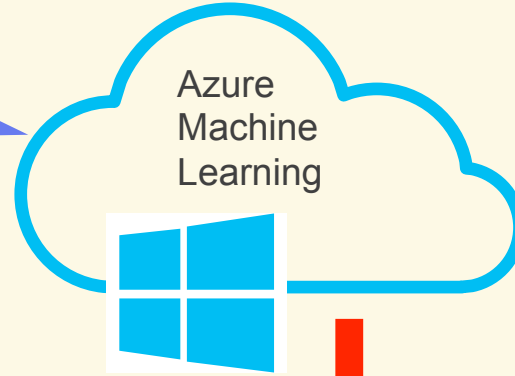




# The Journey: Step 2

Step2: Leverage the most powerful cloud on the planet

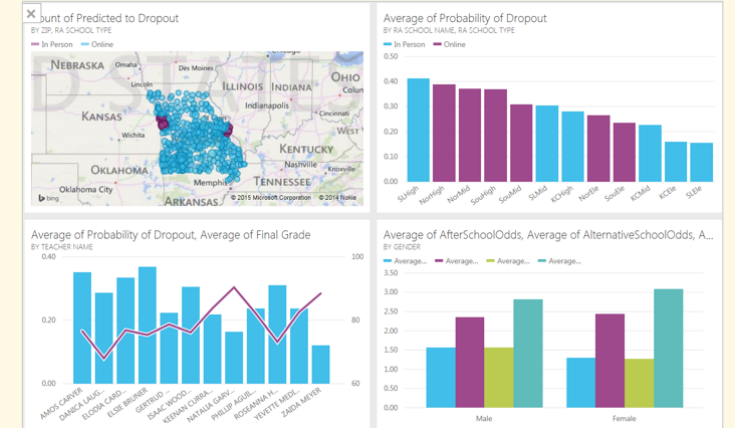
Massive Data: No Problem  
Complexity: No Problem



Data



Predictions



## Student Info

- 10 Variables
- Starting life in teacher grade books

## Azure Machine Learning

- Send to cloud via Excel
- Bring back predictions

## PowerBI

- Self-Serve
- Data Visuals

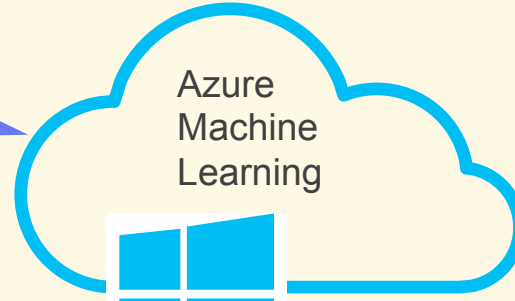


Step1: Start with tool you know well, and data you already have

# The Journey: Step 3

Step2: Leverage the most powerful cloud on the planet

Massive Data: No Problem  
Complexity: No Problem

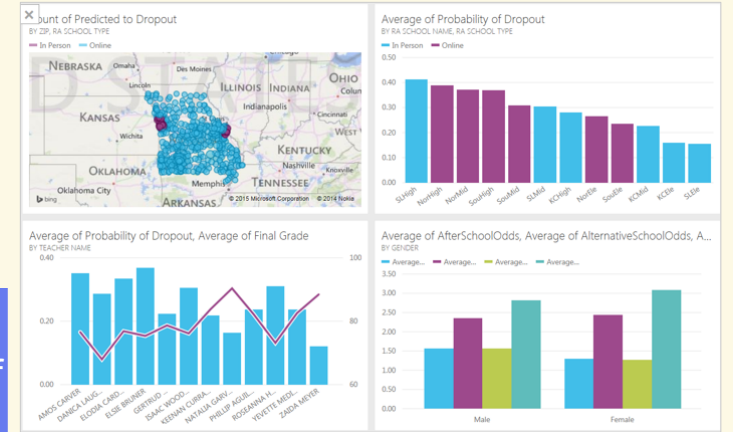


Step3: Leverage Self Service BI

Data



Predictions



## Student Info

- 10 Variables
- Starting life in teacher grade books

## Azure Machine Learning

- Send to cloud via Excel
- Bring back predictions

## PowerBI

- Self-Serve
- Data Visuals



Step1: Start with tool you know well, and data you already have

# Our Results What if scenario

The dashboard displays the following visualizations:

- Count of StudentID and Average of Final Grade...**: A bar chart comparing 'In Person' and 'Online' categories. The 'Online' category shows a significantly higher count (around 4.5K) compared to 'In Person' (around 1.1K).
- StudentID by Race**: A treemap visualization showing the distribution of students by race. The largest segment is 'White (non-Hispanic)', followed by 'Black (non-Hispanic)', 'Hispanic', and 'Asian/Pacific Islander'.
- Count of StudentID and Average of Final Grade by RA School Name and RA...**: A combination bar and line chart showing counts for various RA School Names (KCEle, KCMid, NorHigh, SLEle, SLMid, SouHigh) and their corresponding average final grades. The 'Average of Final Grade' is represented by a red line.
- Table: RA School Name, RA Principal, Average of Final G..., Count**

RA School Name	RA Principal	Average of Final G...	Count
SLEle	Jeff Gutjahr	87.63	
KCEle	Theresa Butts	83.42	
SLMid	Nikki Loyd	82.26	
SouEle	Sandra Werr	80.65	
SouMid	Carrie Bancroft	79.58	
NorEle	Danene Betta	78.38	
KCMid	Anita Nickless	77.03	
NorMid	Marcy Starbuck	75.03	
SouHigh	Rick Matthiesen	74.31	
NorHigh	Sondra Zerow	74.19	
KCHigh	Regina Steinkamp	73.25	
SLHigh	Carrie Read	70.19	
<b>Total</b>		<b>76.71</b>	

**RA School Ty...**

- In Person
- Online

**RA Principal**

- Anita Nickless
- Carrie Bancr...
- Carrie Read
- Danene Betta
- Jeff Gutjahr
- Marcy Starb...

**RA School ...**

- KCEle
- KCHigh
- KCMid
- NorEle
- NorHigh
- NorMid
- SLEle
- SLHigh

**Visualizations**

**Fields**

- Data
- DataLessons

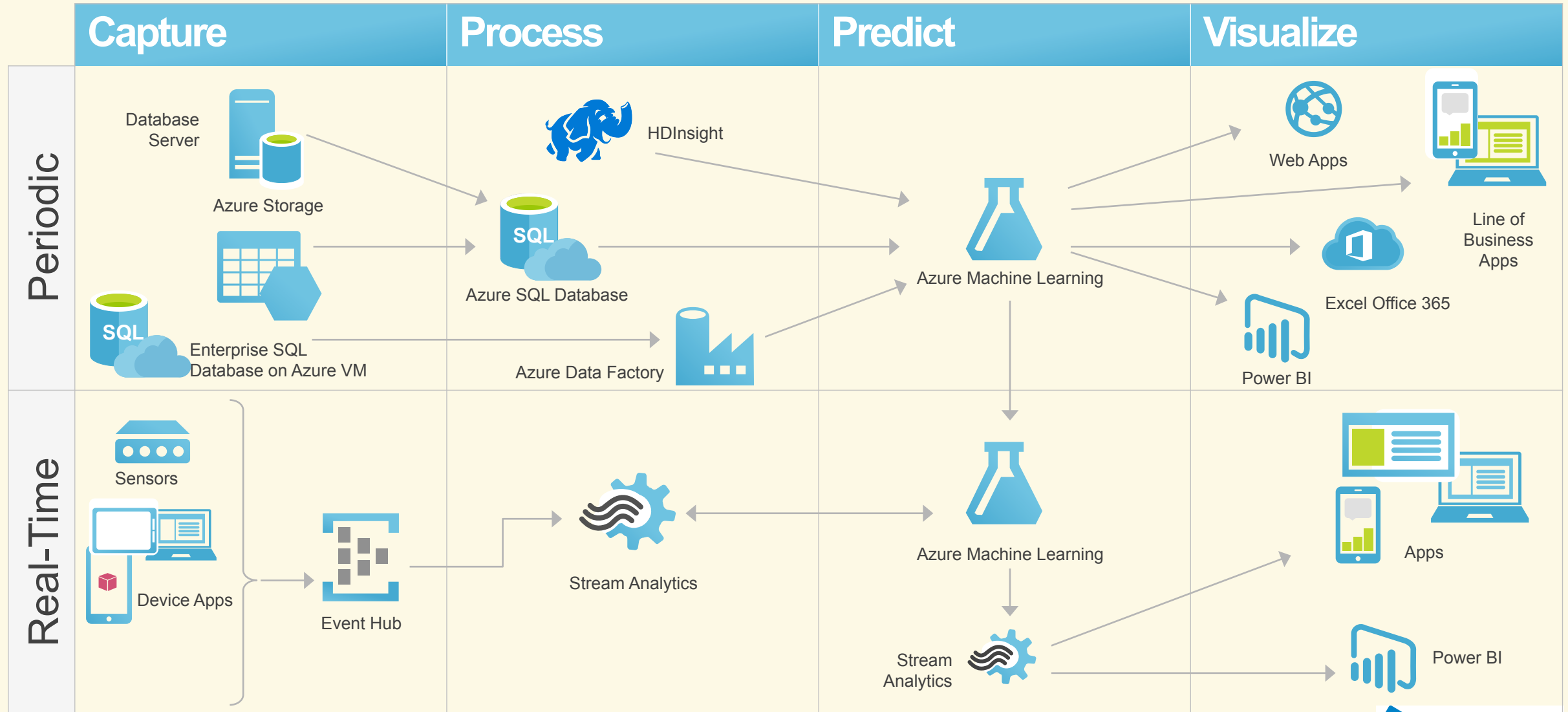
**Filters**

- Page level filters
- Report level filters

**Navigation:** SuperStatus | PrincipalStatus | TeacherStatus | TeacherIntervention | PrincipallIntervention | SuperIntervention

# ***THE TECHNOLOGY***

# Azure Technologies for Predictive Analytics



# GRACIAS!

***Nombre: Jorge Perez Luna***

***Email: Jorge@nealanalytics.com***

***Director Neal México / LATAM***

***Teléfono: +52 55 91316290***

***(Office in Mexico City)|  
Mobile/ 425 698-9831***