Capitalize ANALYTICS

Effectiveness of Machine Learning in K-12 Chromebooks

Introduction - John Crossen

- Various roles working on databases for 20 years
- Data Architect for the past 10 years
- Not a data scientist, but supports them
- Vivit Vertica Special Interest Group (SIG) Leader
- This “LIVE” Sessions is being recorded. Recording are available to all Vivit Members
- If you need technical support, please use the feedback section
- Please submit your questions

Capitalize ANALYTICS

What cool tech was when I was in school…

Capitalize ANALYTICS

What it is now…

Capitalize ANALYTICS

Initial Use Cases

- Machines are shared and stored in each classroom. They do not go home and a student will use a different machine in each class period.
- Do we have enough machines, how can we most effectively distribute machines, and are they being used as intended?

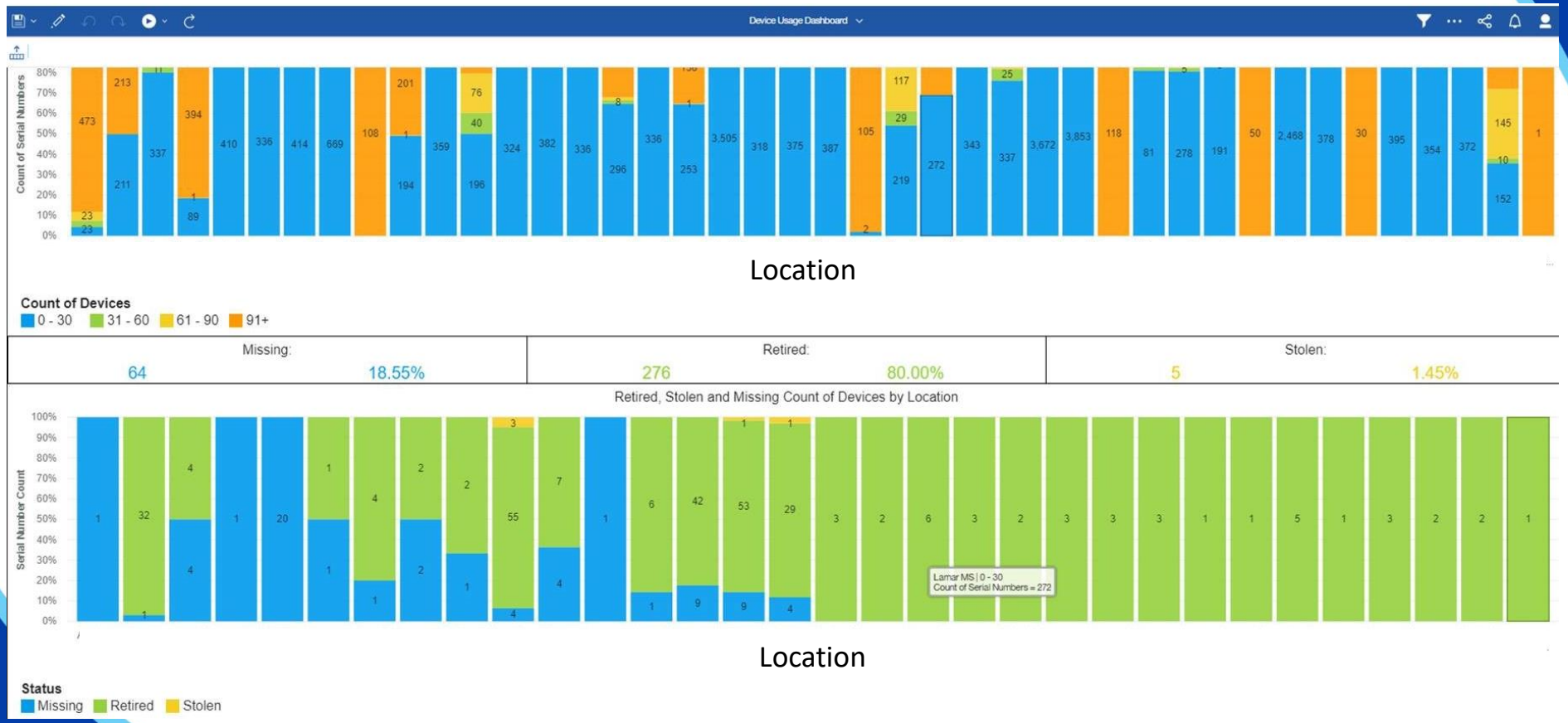
Capitalize ANALYTICS

Project Description

- Tactical goals to start
  - Determine inventory needs.
  - Make sure there are enough machines for each class
  - Prevent over ordering to save money

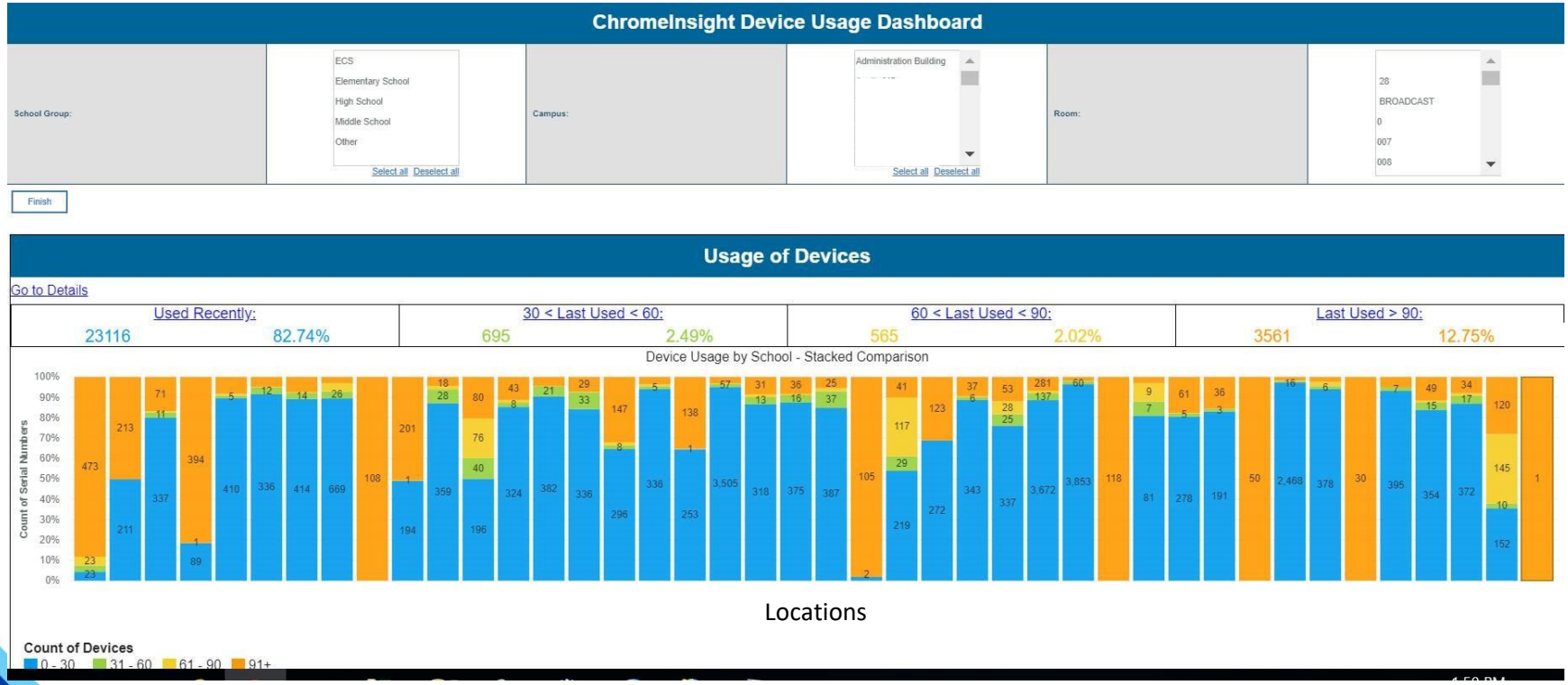
Capitalize ANALYTICS

Device Allocation



Capitalize ANALYTICS

Additional Device Usage



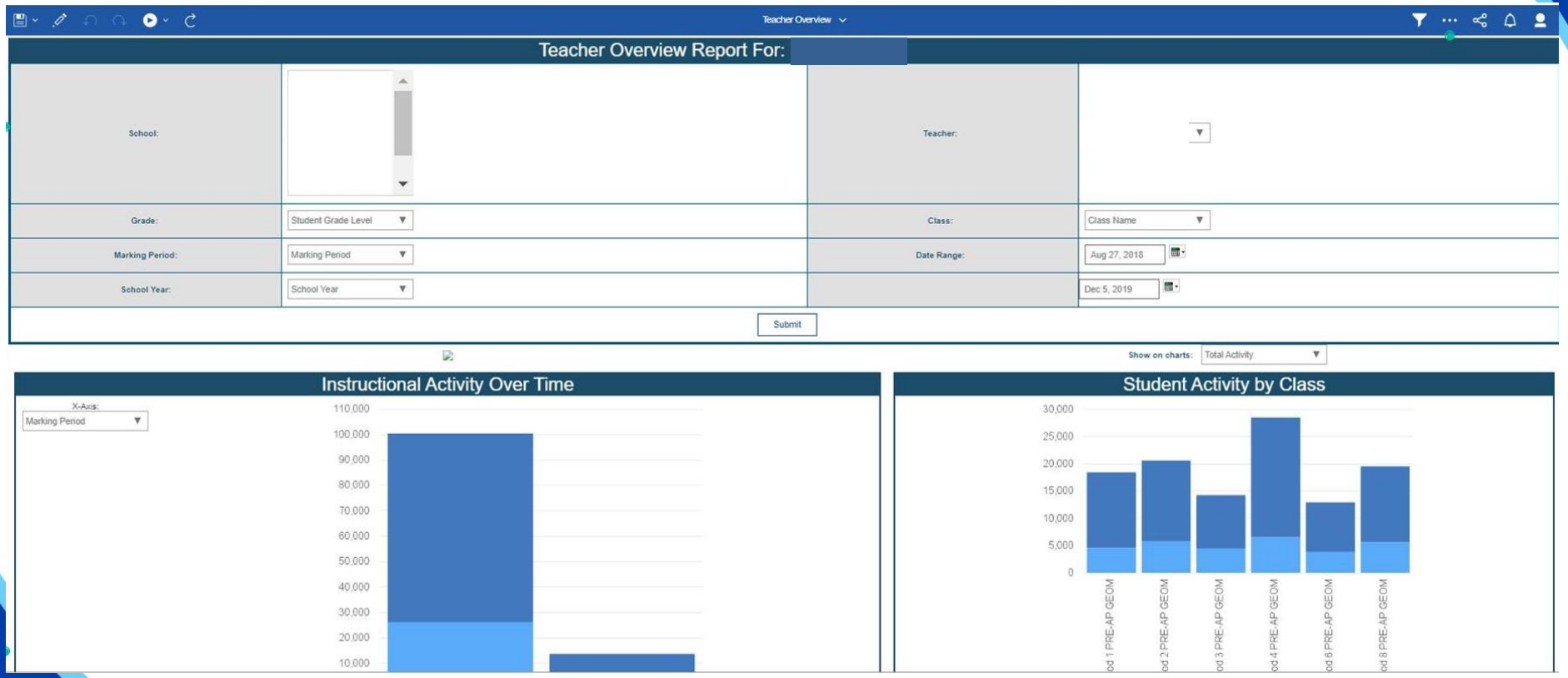
Capitalize ANALYTICS

Tactical Goal 2

- Show what students are doing in a given class.
- What are they clicking on?
- How much time do they spend on a site?

Capitalize ANALYTICS

Device Usage Reporting



Capitalize ANALYTICS

Tactical Goal 3

- Determine usage of paid content.
- Schools purchase licenses for education software, determine how it is used

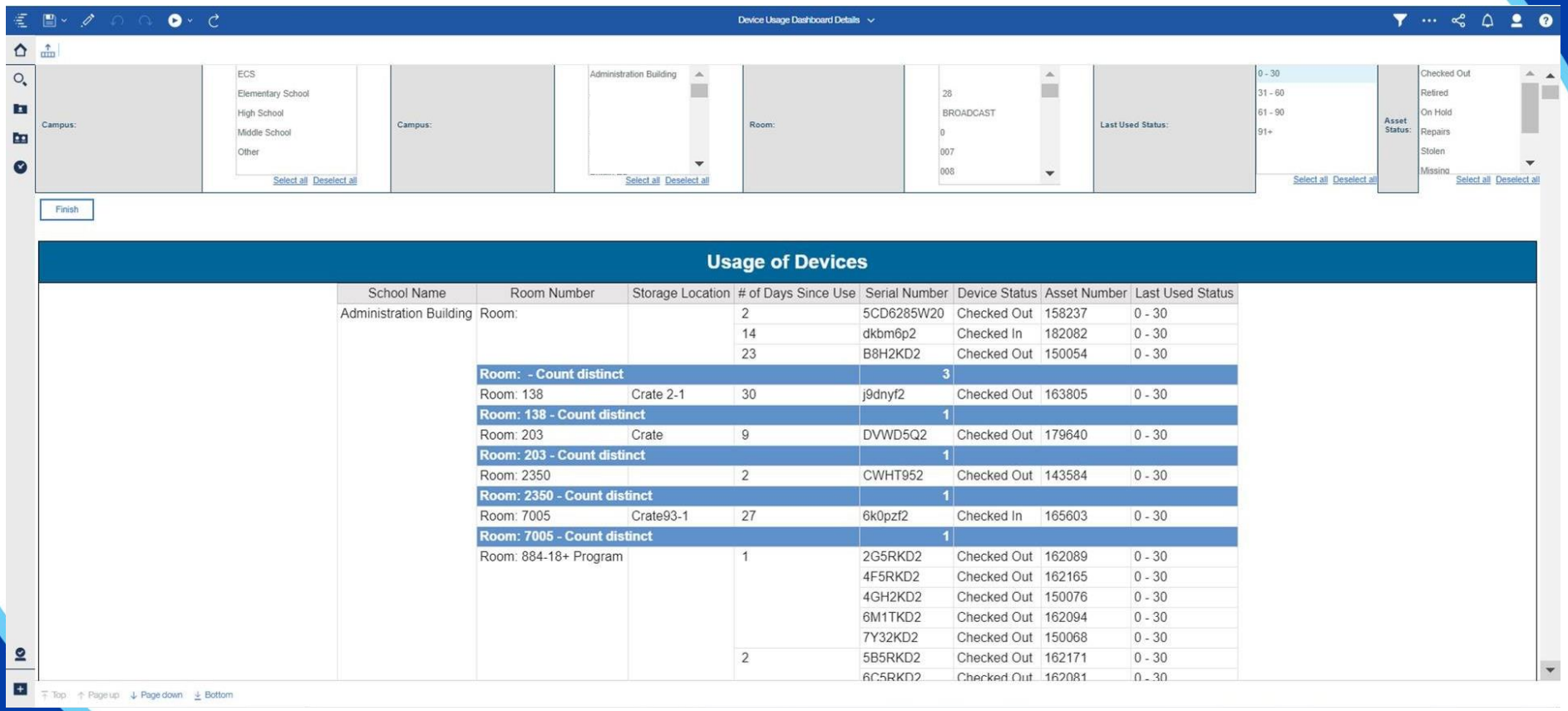
Capitalize ANALYTICS

Tactical Goal 4

- Compare usage across
  - Teachers
  - Departments
  - Classes
- Who is using the tools more

Capitalize ANALYTICS

Usage Statistics



School Name	Room Number	Storage Location	# of Days Since Use	Serial Number	Device Status	Asset Number	Last Used Status
Administration Building	Room:		2	5CD6285W20	Checked Out	158237	0 - 30
			14	dkbm6p2	Checked In	182082	0 - 30
			23	B8H2KD2	Checked Out	150054	0 - 30
	Room: - Count distinct			**3**			
	Room: 138	Crate 2-1	30	j9dnyf2	Checked Out	163805	0 - 30
	Room: 138 - Count distinct			**1**			
	Room: 203	Crate	9	DVWD5Q2	Checked Out	179640	0 - 30
	Room: 203 - Count distinct			**1**			
	Room: 2350		2	CWHT952	Checked Out	143584	0 - 30
	Room: 2350 - Count distinct			**1**			
	Room: 7005	Crate93-1	27	6k0pzf2	Checked In	165603	0 - 30
	Room: 7005 - Count distinct			**1**			
	Room: 884-18+ Program		1	2G5RKD2	Checked Out	162089	0 - 30
				4F5RKD2	Checked Out	162165	0 - 30
				4GH2KD2	Checked Out	150076	0 - 30
				6M1TKD2	Checked Out	162094	0 - 30
				7Y32KD2	Checked Out	150068	0 - 30
			2	5B5RKD2	Checked Out	162171	0 - 30
				6C5RKD2	Checked Out	162081	0 - 30

Capitalize ANALYTICS

Capitalize ANALYTICS

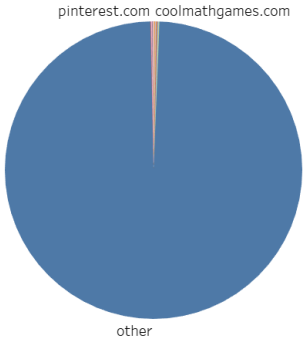
Limits to Reporting

- Many:
  - Domains
  - Students
  - Classes
  - Schools
- Unable to determine trends easily
- Too much data

Capitalize ANALYTICS

Limits to Reporting

Top 20 Pie Chart



Report of Domains

Base Domain	
google.com	97
irvingisd.net	97
youtube.com	93
clever.com	85
quizlet.com	85
achieve3000.com	82
quizizz.com	82
twitter.com	76
wikipedia.org	76
khanacademy.org	75
kahoot.it	74
doubleclick.net	71
explorelearning...	71
pinterest.com	70
weebly.com	70
coolmathgames...	69
desmos.com	69
edgenuity.com	69
amazon.com	67
duckduckgo.com	67
edpuzzle.com	67
mit.edu	66
apple.com	65
forms.gle	65
microsoftonline...	65
minecraft.net	65
schoolobjects.co..	65
co.uk	64
instagram.com	64
opendns.com	64

Capitalize ANALYTICS

Strategic Goals

- Use Machine Learning to determine
  - Patterns
  - Predict usage
  - Effectiveness tied to outcomes (grades)

Capitalize ANALYTICS

More use cases

- How can we see patterns in where students are clicking?
- How do different teachers in the same subject differ in machine usage?
- How frequently is paid content being used?

Capitalize ANALYTICS

Tool Sets

- SQL Server for Standard Reports
  - Pre-existing database
- Combination of Cognos and Tableau for presentation
  - Cognos pre-existing
  - Tableau for newer reports
- Vertica for Machine Learning and Analytics
  - Able to deploy on premise and/or in the Cloud
  - Good selection of Machine Learning functions
  - Fast

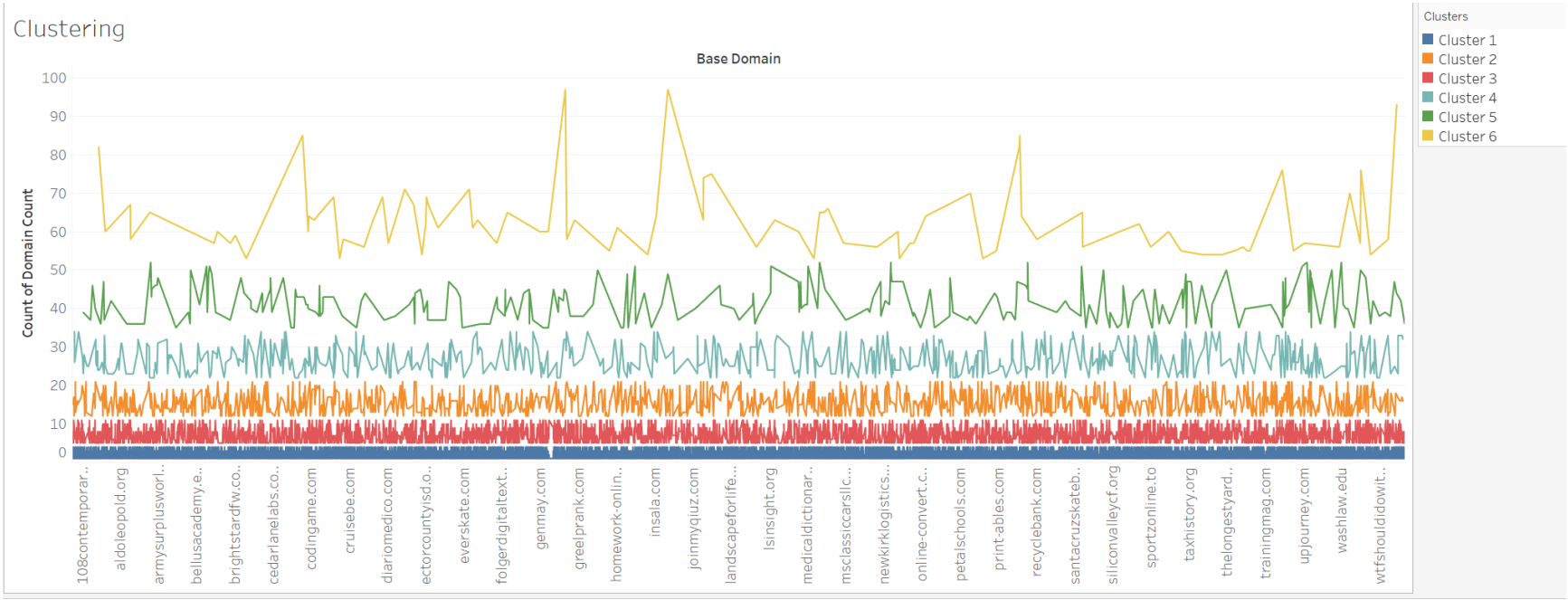
Capitalize ANALYTICS

Domain Clustering

- Need a visual showing domains clustered by us
- Filtered by department
- Filtered by class
- Drill Down on clusters and domains

Capitalize ANALYTICS

Domain Clustering



Capitalize ANALYTICS

Domain Clustering

- Verify underlying data is correct and clean
- Run a model with 1 short statement
- Query runs in 3 seconds over 400M records

```
Select kmeans('public.CIKmeans2', 'demodata_ids', 'domain_id,department_id', 10
USING PARAMETERS exclude_columns='student_rollup_id, class_rollup_id, datekey,
domain_id, domaincount, instructional, paid, exempt, classid, teacherid, school_id',
max_iterations=30, epsilon=0.0001, init_method='kmeanspp',
distance_method='euclidean', output_view='CI_Kmeans',
key_columns='domain_id,department_id');
```

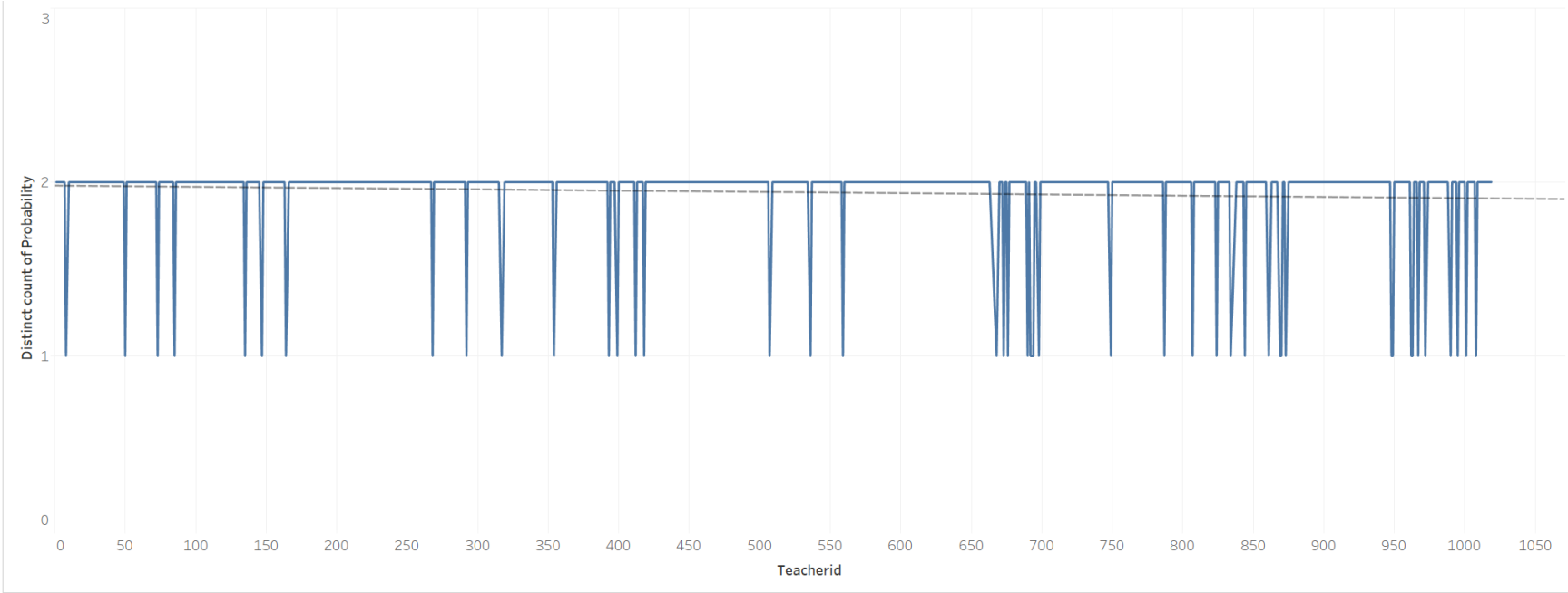
Capitalize ANALYTICS

Linear Regression and RF Clustering

- Show probabilities of students using a particular domain based on the teacher
- Filtered by department
- Filtered by class
- Drill Down on clusters and domains

Capitalize ANALYTICS

Linear Regression and RF Clustering



Capitalize ANALYTICS

Linear Regression and RF Clustering

- Verify underlying data is correct and clean
- Run and combine the results of 3 modes
- Query runs in 40 seconds over 400M records

Capitalize ANALYTICS

Linear Regression and RF Clustering

```
SELECT *,PREDICT_LINEAR_REG(department_id, teacherid USING
PARAMETERS model_name='CILinearReg') into CiLinear FROM
demodata_ids ORDER BY domain_id;

SELECT RF_CLASSIFIER ('CIRF', 'demodata', 'base_domain', '
instructional, exempt' USING PARAMETERS
exclude_columns='base_domain');

SELECT PREDICT_NAIVE_BAYES_CLASSES (teacherid,instructional
USING PARAMETERS model_name = 'naive_CI_model',
key_columns = 'teacherid', exclude_columns = 'teacherid', classes 'list
of class ids' OVER ()
into Naive_Teacher from demodata_ids_distinct;
```

Capitalize ANALYTICS

Future Use Cases

- Tie usage to outcomes (aka grades)
- Create ROI for instructional data
- Incorporate more types of machines/data elements
- Identify teachers using devices most effectively and transfer that knowledge

Capitalize ANALYTICS

Lessons

- “Basic” Reporting is important
  - May show valuable insights
  - Helps raise new questions that can be answered using more advanced techniques
- Must have trust in the data
  - If reports are not right, ML will not be right
  - If reports are not trusted, predictive analytics will not be trusted
  - Trust is hard to gain, easy to lose, very difficult to regain
- Data Modeling is still very important for performance as well as data integrity

Capitalize ANALYTICS

Lessons

- Predictive and ML can gain valuable insights
  - Always take into account the human factor, especially when humans are the subject
  - Promote discussion of solutions, not automate decision making
  - Experiment with inputs and try to find answers even when there is no question
  - Realize these projects are never “done”
  - Have a good framework to add new data sources that ensure trust in the data and maintain good performance

Capitalize ANALYTICS

Questions?

Capitalize ANALYTICS