Driving Value with Big Data

Why Paying Attention to Big Data is so Important for the Public Utility Industry

John Lucker
Principal – Deloitte & Touche LLP

AEGIS *2017* **PHC**

About the Speaker

John Lucker
Principal
Deloitte & Touche LLP
(860) 725-3022
JLucker@deloitte.com
Twitter: @JohnLucker

- Deloitte's Global Advanced Analytics Market Leader and a senior insurance industry strategist, innovator, and analytics evangelist
- Leads delivery of end-to-end consulting services to help insurance and other industry leaders and boards of directors solve and implement solutions for their most complex business issues

AEGIS *2017* **PHC**

About the Speaker

John Lucker
Principal
Deloitte & Touche LLP
(860) 725-3022
JLucker@deloitte.com
Twitter: @JohnLucker

- Expertise spans advanced analytics, product development, pricing / underwriting, claims, distribution, and evolving technologies including IoT, telematics, cyber, risk sensing, robotics, ERM, customer insights, data monetization, behavioral economics and decision bias, and compliance issues surrounding big data

AEGIS *2017* **PHC**

About the Speaker

John Lucker
Principal
Deloitte & Touche LLP
(860) 725-3022
JLucker@deloitte.com
Twitter: @JohnLucker

- Author of more than 90 articles and papers on a variety of advanced analytics, business, and technology topics
- Speaker at numerous global conferences on a variety of analytics and business topics and extensively published in professional journals and publications
- Co-inventor of 5 analytics patents and 2 pending patents
- Education: University of Rochester – B.A.
  University of Rochester – Simon School – M.B.A.

AEGIS *2017* **PHC**

Intro to Big Data and Why 'Value' is the Fourth V

What is Big Data (BD)?

- Some people think of BD as anything that doesn't fit in an Excel spreadsheet!
- In some simple contexts, BD is used to refer to any form of statistical analytics or predictive modeling using diverse internal / external data sources
- In many contexts, BD is shorthand for the granular and numerous data sources used for analytics projects – big, small, old, new, structured, unstructured, etc.

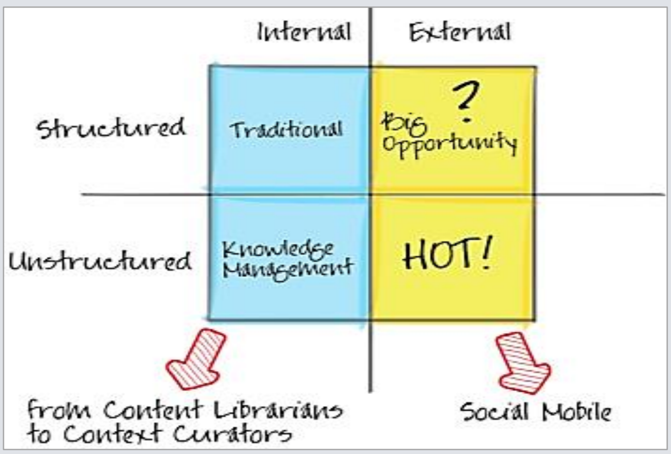
What is Big Data (BD)?

- In more esoteric data contexts, some examples include geospatial, social / sentiment, audio / video, mobile information, telematics, telephonic, internet searching, web logs, etc.

Big Data refers to internal and external data that is multi-structured, generated from diverse sources in real-time and in large volumes, making it beyond the ability of traditional technology to capture, manage, and process within a tolerable amount of elapsed time

Big Data is Constantly Growing, Changing, and Evolving

Are you using all your internal data assets? What external data is relevant for you? Synthetic data?



Common Big Data types include		
Documents (e.g. Medical)	Transactions over time	Surveys & customer feedback
Call / IP Detail Records	RFID	Sensor network
Location based	Web logs	Social Media
Internet Text & Search	Image & photo	Audio
	Video	

Public Utilities Like Natural Gas Companies Possess Diverse and Large Quantities of Data

But How Much?

Internal Data:	External Data:
• Geospatial pipe system layout	• Building priority
• Pipe maintenance & repairs	• Soil composition
• Machinery maintenance & repairs	• Historical temp & weather
• Call & problem reports	• Subway zone
• Historical leak history & locations	• Traffic
• Pressure of pipe	• Flood zone
• Road maintenance & repairs	• Potholes & road repair
• Main replacement prioritization	• Construction
• Asset inspection results	• Customer complaints
• Corrosion history	• Equipment reliability
• Meter inspections	

True Big Data?

AEGIS *2017* **PHC**

And Electric Utilities…

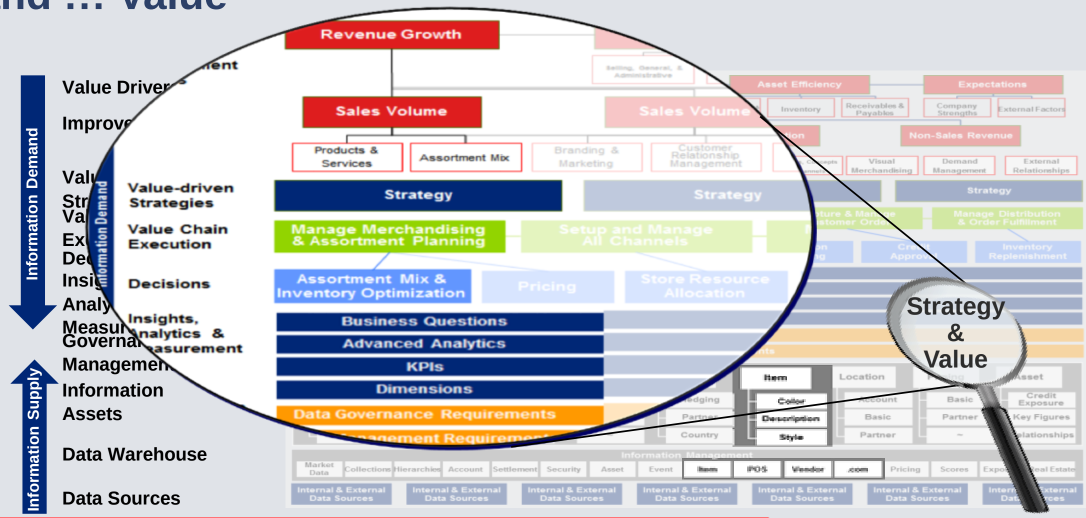
But How Much?

Internal Data:	External Data:
• Geospatial transmission / grid layout	• Building priority
• System history / age & asset composition	• Historical temp & weather
• Pole / wire maintenance & repairs	• Tree / hazard proximity & maintenance
• Equipment maintenance & repairs	• Population density
• Call & problem reports	• CBYD information
• Historical line failure history & locations	• Potholes & road repair
• System / grid stress statistics	• Construction
• Asset / replacement / prioritization	• Customer complaints
• Asset inspection results	• Equipment reliability
• Corrosion / decay history	
• Meter inspections	

True Big Data?

AEGIS *2017* **PHC**

The Three V's and One More – Volume, Velocity, Variety, and … Value



The Evolution of Big Data & Some Issues Companies & Boards Must Pay Attention To

While big data has driven enormous focus and organizational spending, hype should not drive the attention away from these issues:

Data Quality

Data cleaning might take more effort than data analysis. The magnitude of bad data will get amplified with the increase in type of sources, and volume of data – but keep 80:20 Rule in mind

Relevance

The amount of data stored will have contextual relevance with individuals analyzing the data.

'One Man's Signal is another Man's Noise'

Interpretability

Substantial amount of data is unstructured and hence liable to different interpretation by different people and machines

Business decisions stand the risk of being based on biased interpretations

Privacy Issues

As organizations capture more volumes of data from various sources, they are more susceptible to disturbing privacy concerns

Especially as more and more consumer data is being used, organizations will have to be sensitive about the data they use

Redundancy

There are chances, organizations are capturing same data from multiple sources, multiple times

e.g., Tweets, updates

Organizations should be wary they are not investing in capturing one data point from multiple sources

Novelty

Most of the time, a lot of data captured from Big Data sources is already captured in existing data available with the enterprise

Big Data investments should focus on finding new insights

Dilutes Value Focus

With Big Data hype, a lot of attention is going into collection, storage, and access of Big Data

This has diverted attention from analysis and ultimate use of data

Knowing what organizations want to do with the data might also be an important question to consider

Avoid "Hoarding"

Complex

Big Data is still confusing to many professionals

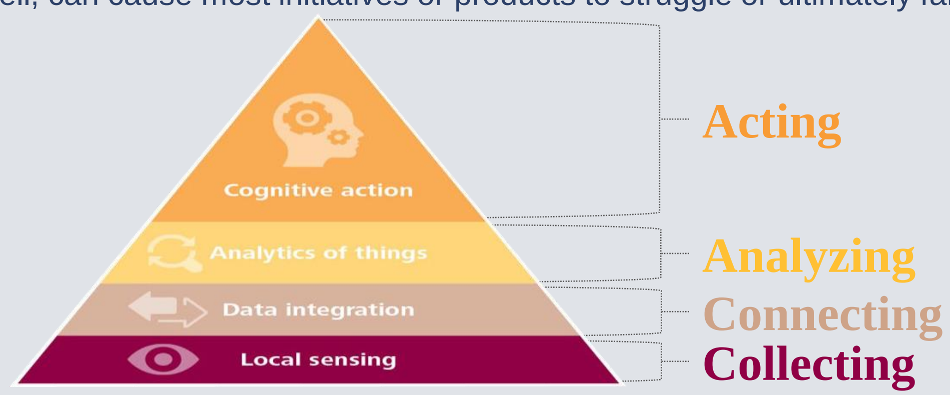
After the hype subsides, efforts into making it less complex, and user friendly should ensue

Source: British Retail Consortium (BRC) Analysis

Some well known and evolving examples of Big Data and how it's used in groundbreaking ways

Big Data Technologies Follow a Maturity Hierarchy of Capability and Activity

The bottom layer of activity dominates the discussion among technologists, executives, and boards. However, it is these upper layers of activity that, if not done well, can cause most initiatives or products to struggle or ultimately fail.



AEGIS 2017 PHC

Case Study 1 – Telematics Program

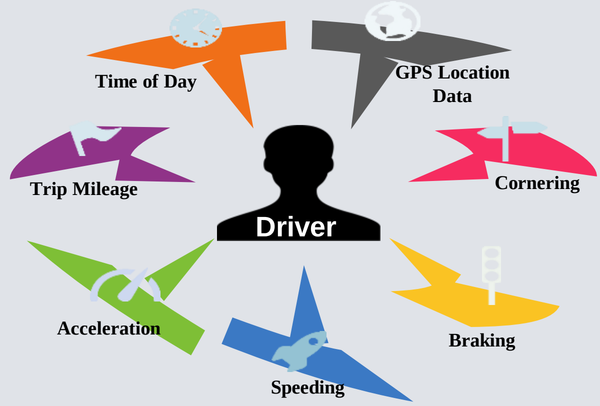
Commercial auto

What Policyholders See — Mobile App

What the Company Sees — Insurer Portal

- Aggregated driving behavior
- Individual trip detail
- Automatic trip detection
- Real-time driver feedback
- Individual trip score
- Driver behavior fingerprint

- Advanced analytics and reporting
- Quantitative driver scores
- Configurable scoring algorithms and branding
- Customer support



1. Summary data
2. Event data
3. Stream data
4. Background data

AEGIS 2017 PHC

Case Study 2 – Wearables – Employee Wellness, Workplace Safety, & Workers' Comp

Anomaly Detection

Hit 10,000 steps today?
Pulse rate is outside normal range for age and exertion level

Comparative Usage

Your calories burned vs. others
Workers in job class X at Y age lift Z / day

Optimization

When in your sleep cycle to wake up
You have exerted yourself enough today

Prescription

Time to get off the couch!
Don't lift that – too heavy!
Attach your safety harness!

AEGIS 2017 PHC

Some risks of Big Data, the analytic ecosystem, and related technology

Companies Struggle to Understand Big Data Value, so They Tend to Hoard

- Companies are aggregating vast amounts of disparate data into data lakes, data marts, and data warehouses
- The costs for aggregation, storage, and maintenance are significant
- The value of these data stores is not yet clear – companies “don’t know what they don’t know” so they err on saving everything without a clear plan

Source: http://all-that-is-interesting.com/andy-warhol-hoarder

AEGIS 2017 PHC

Companies Struggle to Understand Big Data Value, so They Tend to Hoard

- Robust risk management of these assets is essential – security, cyber, analytics, model risk, etc.
- Strategic execution required to obtain latent or organic value

Source: http://all-that-is-interesting.com/andy-warhol-hoarder

AEGIS *2017* **PHC**

Are There Accuracy Issues with Some Big Data?

A sample of people surveyed reviewed their personal data from a major national consumer data broker and reported that nearly half of the variables about them were a coin flip or less in terms of correctness!

The average data accuracy rating was about

50%

and typically worse if a participant was born outside the US or had moved to this country within the past few years despite having a significant financial footprint in the US (apartment rental, home / car ownership, credit cards, established US career, etc.)

Big implications for micro vs macro analysis

Source: https://www.linkedin.com/pulse/20141118145642-24928192-predictably-inaccurate-big-data-brokers

AEGIS *2017* **PHC**

Failure Risks in the Big Data and IoT Ecosystem

Constant firmware updates that break some features

IoT equipment sensor malfunctioned in the winter causing non-detection of a hazardous condition

RISK

"The app logged itself out on my truck tablet so I didn't get the alert about the equipment failure"

If a vehicle crash is due to programming error, the error could impact many vehicles at the same time

Brand X fitness band performed best with a x.x% error rate. Brands A, B, C, D, E had error rates of a.a%, b.b%, c.c% - worst was brand M with d.d%

AEGIS *2017* **PHC**

Case Study of Utility Asset & Distribution Risk Analytics

Context – Understanding the Challenges of Utility Asset & Distribution Risk

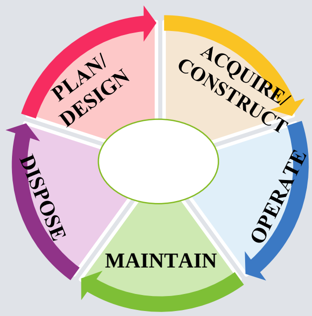
How do I ensure **capital allocation** is made in a consistent, transparent and optimal way across all assets?

How do I run altern**ative planning scenarios** for my investment decisions?

What is the **life expectancy** of an asset's component or part?

How can I optimize my **investment** and **operational and maintenance** spend?

What **impact** will my investments have on the business' **strategic objectives**?



How can I perform in depth **root cause** failure analysis on my processes and assets?

How to optimize my **whole-life-cycle costs** for my assets?

What are the major **risks** impacting asset and process performance?

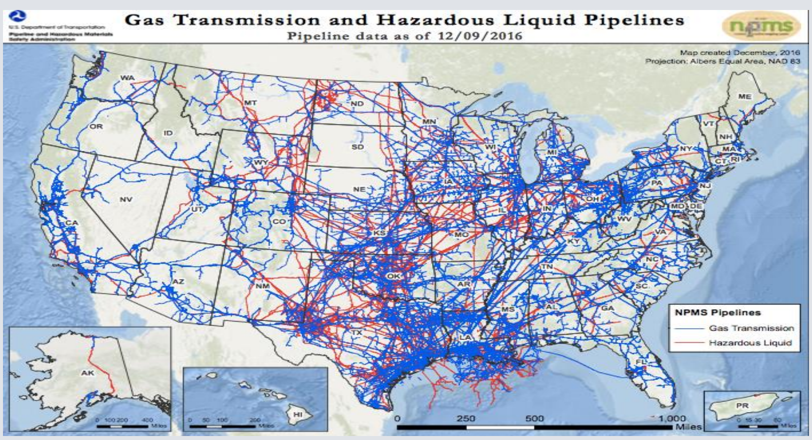
How can I **monitor** and optimize asset and process performance?

How can I detect issues sooner and intervene **pro-actively**?

How can I achieve optimal asset **efficiency** and **availability**?

AEGIS 2017 PHC

The US National Gas Transmission System is Huge and Complex



Source: https://www.npms.phmsa.dot.gov/Documents/NPMS_Pipelines_Map.pdf

AEGIS 2017 PHC

At An Individual Regional or Local Utility Level, Even More Complexity Exists

- At regional / local levels, gas transmission / distribution networks are stressed due to aging infrastructure, community growth & expansion, and delayed asset investment
- Over the next 20 years, the utility industry expects to spend over $2 trillion on infrastructure remediation for gas, electricity, and water infrastructures
- Infrastructure remediation partially driven by public awareness of San Bruno and NYC explosions, LA gas leaks, Flint water crisis, and electric grid blackouts

AEGIS 2017 PHC

At An Individual Regional or Local Utility Level, Even More Complexity Exists

- Heightened regulatory scrutiny at federal, state, local levels – increased focus on data gathering, proactive analytics, and risk-based data-driven decision processes
- Regulators aren't just focused on long-tail management but rather are expecting short / medium term incremental improvements and risk based approaches

AEGIS 2017 PHC

Project Objectives

Proactively identify specific asset components that need maintenance or replacement to prevent catastrophic failures

- Incorporate deep data insights, looking at the effects of multiple variables on asset health and safety
- Identify risks and priority, based on existence and location of risk factors correlated with failures.

Reduce frequency and consequence of system failures including loss of service, loss of life or injuries, property damage and environmental and reputational impacts

- Take immediate action, based on the correlative single and multivariate analysis and insights that emerge
- Use data-driven strategy to drive budgets and approvals, both inside the company as well as with regulatory agencies
- Protect employees, the community, and the company

Improve capital and operational maintenance

- More efficient capital / operating budget usage through risk-based asset replacement and repair
- Improve regulatory reporting capabilities to FERC, PHMSA, and state PUCs
- Better deployment and utilization of limited replacement and repair personnel
- Achieve short / medium term remediation while long term sensor systems installation in progress for many years

Types of Data Sources – Internal

- Geospatial system layout (lat / long)
- Pipe maintenance & repairs
- Machinery maintenance & repairs
- Call & problem / incident reports
- Historical leak history & locations
- Pressure of pipe – time based
- Burial maintenance & repairs
- Main replacement prioritization
- Asset inspection results
- Corrosion history
- Meter readings & inspections
- Pipe materials & characteristics
- Pipe installation characteristics
- Pipe fittings, materials, quantity
- Customer usage & priority
- Historical component defects
- Cave-ins and road failures
- Historical installer info & history
- Cathodic protection

Types of Data Sources – External

- Building usage & priority
- Soil composition & history
- Historical temp & weather
- Permits & regulatory info
- Subway zone
- Traffic volumes
- Wetlands & flood zones
- Potholes & road repair
- Construction projects
- Customer complaints
- Equipment reliability

AEGIS *2017* **PHC**

Types of Data Sources – Synthetic

- Various norms and statistics for assets with each geospatial location or component measured relative to population
- Climate variability and soil volatility variables
- Proximity to assets categorized as cautionary / abandoned / removed from system
- Valve count relativities
- Corrosion proximities (relative to other non system assets)
- Proximal repairs relative to each geospatial location
- Time since various inspection actions

AEGIS *2017* **PHC**

Exploratory Data Analysis and Data History Research

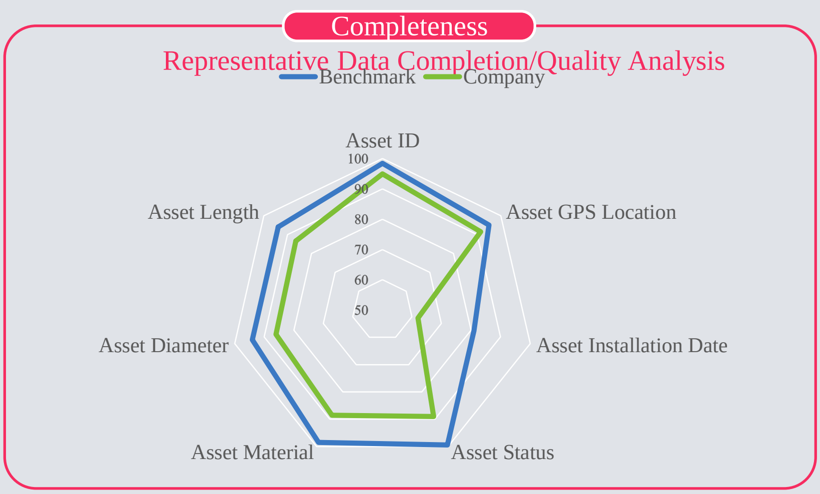
Representative tasks:

- Identify system field usage history
- Study distributions and values for all key variables and ensure what you see matches documentation and metadata
- Identify duplicative variables and identify source of primary info and best use
- Study gaps in geospatial information and assess geospatial accuracy
- Study distribution of leaks to determine leak / repair frequency / severity trends
- Study data to discard and modeling ramifications of non-use

AEGIS *2017* **PHC**

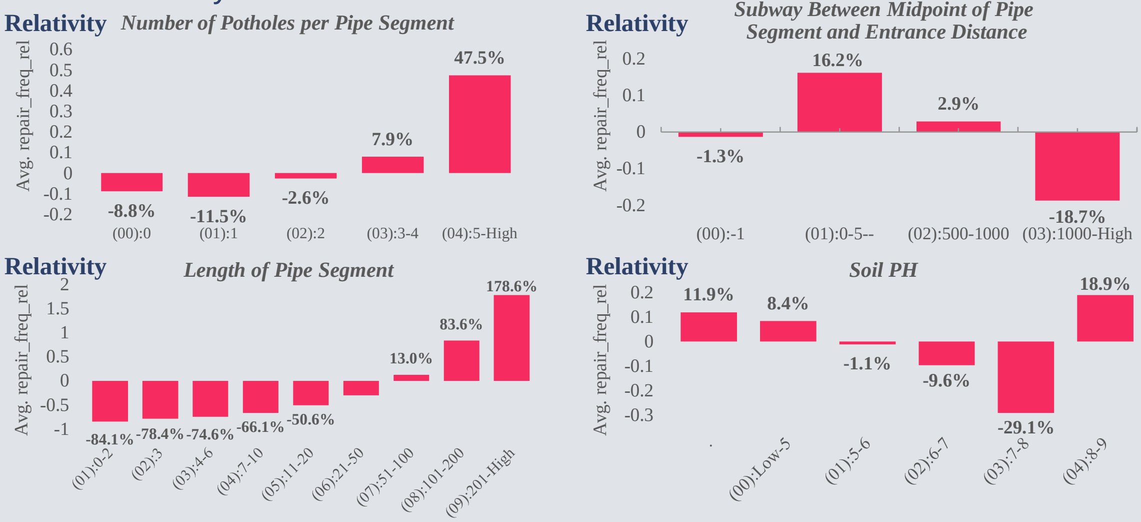
Exploratory Data Analysis and Data History Research

	Benchmark	Company
Asset ID	98.5%	95.0%
Asset GPS Location	95.0%	91.5%
Asset Installation Date	81.0%	62.0%
Asset Status	99.5%	89.0%
Asset Material	98.5%	88.5%
Asset Diameter	94.0%	86.0%
Asset Length	94.0%	86.5%



AEGIS *2017* **PHC**

Predictive Asset Management
Univariate Analysis



AEGIS *2017* **PHC**

Predictive Asset Management
Multivariate predictive variables

Variable	Predictive Direction	Description
CONSTRUCTION	↑	Pipe is near construction
CAR_LEAK_COUNT	↑	Historical number of Cave-ins resulting in leaks of pipe
PAVING_LEAKS_COUNT	↑	Historical number of paving leaks of pipe
PIPE_AGE	↑	Age of pipe
CARSCOUNT	↓	Historical number of cave-ins near pipe
FITTINGSCOUNT	↑	Number of fittings of pipe
MEASUREDLENGTH	↑	Measured length of pipe
PRESSURE_NORM	↑	Pressure of pipe divided by length of pipe
MATERIAL_S	↑	Pipe material is steel
MATERIAL_C	↑	Pipe material is cast iron
MATERIAL_W	↑	Pipe material is wrought iron
MATERIAL_P	↓	Pipe material is plastic
NO_CLAMPS	↑	Number of clamps of pipe
CQF	↓	Main Replacement Prioritization (MRP) – Consequence Factor (CQF)
LRL_COUNT_PREV1	↑	A leak repair was performed on the pipe the previous year
PRECIPITATION_PREV1	↓	Amount of precipitation in prior year
TEMPERATURE_PREV1	↓	Temperature of prior year
CV_MAIN_PRESSURE_PROX_100	↓	Coefficient of variation of main pipe pressure for all pipes within a proximity of 100 ft
SUBW_ENT_DIST	↓	Distance (in feet) between the midpoint of the pipe segment and a subway entrance
LEAK_ACT_SURVEILLANCE100_PREV1	↑	Leaks inspections within 100 ft of the pipe

Key

↑ As variable increases, all else being equal, the probability of leak increases

↓ As variable increases, all else being equal, the probability of leak decreases

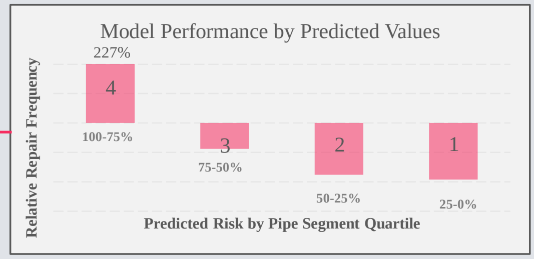
AEGIS *2017* **PHC**

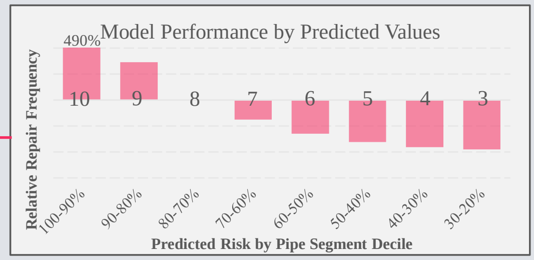
Predictive Asset Management

Model performance lifts

- For the worst 1% of assets, the model identified the population of pipes that were repaired over 1,000% more frequently than average
- For the worst 10% of assets, the model was able to correctly identify pipe segments that needed repair 490% more frequently than average

00%	-	25%	(96%)
25%	-	50%	(88%)
50%	-	75%	(44%)
75%	-	100%	227%
00%	-	10%	(97%)
10%	-	20%	(96%)
20%	-	30%	(95%)
30%	-	40%	(91%)
40%	-	50%	(81%)
50%	-	60%	(65%)
60%	-	70%	(38%)
70%	-	80%	(0.8%)
80%	-	90%	73%
90%	-	100%	490%
90%	-	95%	223%
95%	-	100%	757%
95%	-	96%	391%
96%	-	97%	$468%
97%	-	98%	617%
98%	-	99%	938%
99%	-	100%	1E3%
Total			0.0%

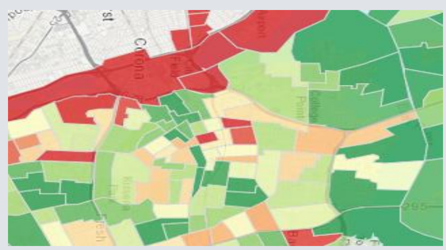




AEGIS *2017* **PHC**

Asset Risk Model Implementation and Usage

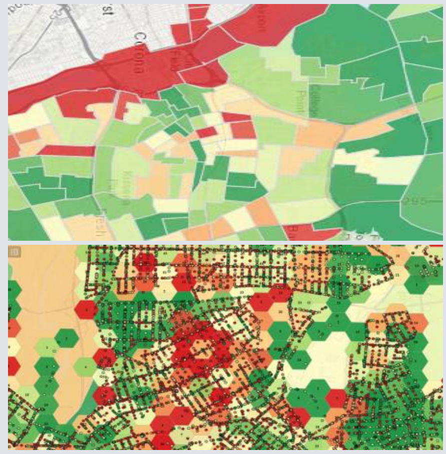
- Predictive model scoring engine generates risk score for individual assets and geospatial points
- Business rules translate scores and reason codes for recommendations for action and reason codes provide problem clues
- Reprioritize asset inspections and repairs – the model better predicts repair needs
- At first, focus on the “worst of the worst” – the model ”reshuffles the deck”



AEGIS *2017* **PHC**

Asset Risk Model Implementation and Usage

- Some correlations and reason codes indicate special remediation, repairs, situational conditions (water, pH, traffic, etc.)
- Visualizations can depict the composite risk score calculated using the predictive model
- Process should ensure gathering of accurate / complete and new / enhanced data going forward



AEGIS *2017* **PHC**

Q&A – and Some Available Reading Material and Much More Upon Request

Email me at JLucker@deloitte.com and I will send you more reading on Big Data

AEGIS *2017* **PHC**

AEGIS

AEGIS 2017 PHC