Innovative approaches to capacity building in the use of Big Data

International Conference on Big Data for Official Statistics
31st August – September 1st 2016
Many claim to be at the forefront of Analytics but how much “intelligence” are you digging up from your data

Pushing the Analytics Envelop

- **Data Queries**
  - What happened yesterday?
  - OLAP (search query, Drill-down Analysis)

- **Descriptive Analytics**
  - Why did it happen yesterday?
  - (correlation, association, factor analysis)

- **Business Intelligence**
  - What is happening now?
  - (dashboard, scorecards, Self-Service BI)

- **Predictive Analytics**
  - What will happen tomorrow?
  - (Classification, Forecasting)

- **Prescriptive Analytics**
  - How to Influence tomorrow outcomes?
  - (Uplift Modeling, Optimization)

**Business Value and Decision-Support**

Low to High
Many claim to be at the forefront of Analytics, but how much “intelligence” are you digging up from your data

Pushing the Analytics Envelop

- **Data Queries**
  - What happened yesterday?
  - OLAP (search query, Drill-down Analysis)
- **Descriptive Analytics**
  - Why did it happen yesterday?
  - Correlation, association, factor analysis
- **Business Intelligence**
  - What is happening now?
  - Dashboard, scorecards, Self-Service BI
- **Predictive Analytics**
  - What will happen tomorrow?
  - Classification, Forecasting
- **Prescriptive Analytics**
  - How to influence tomorrow outcomes?
  - Uplift modeling, Optimization

**Description**

- Reporting
- Discovery
- Monitoring
- Prediction
- Prescription

**Low**

- Data Queries
- Descriptive Analytics
- Business Intelligence
- Predictive Analytics
- Prescriptive Analytics

**High**

- Analytics Governance
- Data Governance
- Technology
- Culture and Leadership
- Analytics Competency
For a start, organizations need to go through a diagnostics heath check in order to assess its analytics fitness.

### Analytics Health Check

#### Stage 1: Analytically Impaired
- Lack of analytics
- Inconsistent data, poor data quality, or poorly-organized data
- No clear procedure for data access and dissemination of information
- Lack of technology
- No awareness or interest
- Few analytical skills

#### Stage 2: Localized Analytics
- Analytics focus is dispersed across multiple targets, which may not be strategically important
- Useable data organized in functional or process silos
- Informal understanding of data usage and access policies
- Data technology and expertise in disparate clusters across organizations
- Interest exists only at the functional or process level
- Isolated pockets of analysts with little communication

#### Stage 3: Analytical Aspirations
- Analytical efforts coalesce around a set of unified targets
- Organization has nascent centralized data repository
- Inconsistent but existing documentation of policies and standards
- Early stages of an enterprise-wide approach
- Leaders recognize the importance of analytics
- Analysts clustered in key target areas

#### Stage 4: Analytical Adopters
- Analytical activity centered around defined key domains
- Integrated and accurate central data warehouse
- Clear and complete documentation of data policies, e.g. data privacy, access, etc.
- Key data, technology and analytics are centralized or networked
- Leadership provides support for analytics
- Highly capable analysts in control or networked organization

#### Stage 5: Analytical Frontrunners
- Analytics support the firm’s and strategy and further the firm’s mission
- Relentless search for new data and metrics
- Automated data access procedures
- All key analytical resources centrally managed
- Leadership actively encourages and supports analytics and exploration
- Professional analysts with strong training and support for amateur analysts

---

**Where is your organization today?**
And that needs to take place across the main facets of analytics, beginning with technology and ending with people competency.
Mapping analytical goals to technical requirements requires diverse knowledge of technology architecture and scalability.

<table>
<thead>
<tr>
<th>Source Data</th>
<th>Store Data</th>
<th>Convert &amp; ETL</th>
<th>Transform Data</th>
<th>Exploratory Analysis</th>
<th>Model Build &amp; Generate Insights</th>
<th>Visualisation</th>
<th>Model Execution in Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Store, Transform and ETL</td>
<td>Data Exploration and Cleansing</td>
<td>Predictive Model Building</td>
<td>Visualization and Presentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The intricacies of Governance extends to everyone within the organization and across the knowledge discovery journey.
“To Govern with Data Requires That You First Govern the Data”

Amjad Zaim, Ph.D
CEO, Cognitro Analytics

"With advancement in big data computing, machine learning and IOT, technology is no longer a hard problem in the game of analytics, governance is !!! To govern with data requires that you first govern the data, that is, bringing the right data products to the right people at the right time and within the right legal and ethical framework"
Cultural Change Requires the Power of Indirect, Simplicity and Universality

“If even if it is analytics, don’t call it analytics”

“What is complex today, needs to be simple tomorrow”

“Single rules”
Single source of truth, single language, single process, single...

Return on Investment
Return on Data
Return on Insight
Return on Innovation
Cognitro analytics capacity building takes an incremental approach of “Engage, Empower, Enable, Educate and Energize”
Cognitro analytics capacity building takes an incremental approach of “Engage, Empower, Enable, Educate and Energize”
Tailored analytics education program needs to examine current capacity and answer key questions across the target audience.
Analytics capacity building programs also need to combine both technical and non-technical courses and for various levels.

### Cognitro Analytics Standard Data Science Training Program

<table>
<thead>
<tr>
<th>Beginners Level Track</th>
<th>Intermediate Level Track</th>
<th>Advanced Level Track</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Learning Goals</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Tech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Navigate the data dimensional space and understand distribution</td>
<td>• Develop predictive modeling</td>
<td>• Master main optimization schemes</td>
</tr>
<tr>
<td>• Develop basic analytics modeling techniques to uncover root causes</td>
<td>• Apply machine learning algorithms</td>
<td>• Understand text mining basics</td>
</tr>
<tr>
<td>• Learn best practices for tackling a business analytics problem</td>
<td>• Develop social network analytics</td>
<td>• Conduct social sensing and sentiment analytics</td>
</tr>
<tr>
<td><strong>Sample Courses</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Fundamentals of Data Science</td>
<td>1. Statistical-Based Algorithms</td>
<td>1. Natural Language Processing</td>
</tr>
<tr>
<td>2. Data Quality Management</td>
<td>2. Learning-Based Algorithms</td>
<td>2. Simulation and Linear Programming</td>
</tr>
<tr>
<td>3. Essential Data Mining Tasks and Algorithms</td>
<td>3. Social Graph Theory</td>
<td>3. Fuzzy Logic and Rule Sets Optimization</td>
</tr>
<tr>
<td>4. Big data and Hadoop Technology</td>
<td>4. Business Intelligence &amp; Data Visualization</td>
<td>4. Real-time Analytics for IOT data</td>
</tr>
</tbody>
</table>

### Data

- Structured data
- Structured and Big Data
- Unstructured text and Big Data

### Training & Scaling

- **Executive Track**
  - Executive workshop

- **Technical Track**
  - Online Courses
  - Classroom-Based Courses

**International Conference on Big Data for Official Statistics**
Starting with the science behind the analytics, or the core data science competency

Cognitro Analytics Capacity Building Framework

- **Define the Problem**
  - Understand Strategy/ Business
  - Understand Data
  - Setting Up the Analytics Environment

- **Design and Develop Analytics Model**
  - Select algorithm and parameters and run model
  - Analyze and interpret results
  - Back-test Model
  - Test Scalability
  - Evaluate Model Fit
  - Define computed variables
  - Develop Data Frame with input features
  - Divide Data into Test and Train Sets

- **Generating and Communicate Insights**
  - Capturing, packaging and promotion
  - Maintenance
  - Storytelling

- **Recommended Analytics Training Framework**
  - Starting with the science behind the analytics, or the core data science competency
  - Learning how to build a predictive model
  - Practitioners
  - Learning how to set up and manage project
  - Learn how to build analytics Organization

Algorithms:
- Regression Algorithm
- Bayesian Algorithm
- Random Forest Algorithm

International Conference on Big Data for Official Statistics
In addition to developing the management skills needed to drive a Big Data project,

Cognitro Analytics Capacity Building Framework

A. PROJECT MANAGER
- Understand the industry and business challenges
- Coordinates requirements gathering
- Locates and obtains required data

B. TECHNICAL LEAD
- Quality and timeliness of deliverables
- Technical direction and "daily stand-up" review

C. SUBJECT MATTER EXPERTS
- Brings specialized tradecraft and best practices in a particular subject area

D. DATA SCIENCE TEAM
- Brings specialized analytics tradecraft and best practices in customer analytics functions
- Examples: Machine Learning Scientist, Natural Language Processing Scientist, and Social Media Engineer

Define the Problem
- Understand Strategy/ Business
- Understand Data
- Setting Up the Analytics Environment

Generating and Communicates Insights
- Capturing, packaging, and promotion
- Maintenance
- Storytelling
And ending with learning ways to nurture and grow an analytics culture with data-driven philosophy

Learning Outcome of Our Customized Program

- Learn how to build a predictive model
- Learn how to setup and manage project
- Learn how to build analytics Organization

### Capability Guide Highlights

- Developed by Cognitro Analytics data science subject matter experts
- How to get buy-in for data science throughout the entire organization.
- Where to place your data science teams in your organization. Should they be centralized, Dispersed or Permanently embedded in individual business units?
- How to stand up the position of Chief Data Officer (CDO).
- How you can leverage our Data Science Talent Management Model.
- Why design thinking—when applied to data science—can unlock organizational value.
Analytics capacity through an incremental approach of “Engage, Empower, Enable, Educate and Energize”

**Igniting Data Science Talent**

<table>
<thead>
<tr>
<th>Capability Assessment</th>
<th>Executive Workshop</th>
<th>Training &amp; Scaling</th>
<th>Capstone Project</th>
<th>Competency Model</th>
</tr>
</thead>
</table>

**Cognitro Analytics Capacity Building Approach**

1. **Executive Workshop (1 days)**
   - Demystify analytics to senior leadership
   - Identify existing analytics gaps within organization
   - Recommend courses & pre-select target groups

2. **Pre-Training (5 days)**
   - Development & delivery of course pack including training goals & objectives
   - Conduct survey-based prequalification
   - Highlight Data Science career outlook

3. **Practitioner Training (15 days)**
   - Build the theoretical background of analytics
   - Demonstrate the analytics value through 3 use cases
   - Conduct hands-on sessions to facilitate “learning-by-doing”

4. **On-the-Job Coaching (10 days)**
   - Develop a capstone project as an analytics exercise
   - Provide on-site supervision on applying gained skills
   - Assess the learning outcome & recommend enhancement
A Competency Model needs to close the loop from, executives level training to on-the-job coaching of data science practitioners.