* + Domain expertise is required
  + Data Scientist, business analyst and software engineer need to work together
  + Data engineer does the data cleansing, wrangling and enrichment
  + Data Scientist does the modeling
  + Business analyst/architect and data scientist provide the enrichment process
  + Data Scientist provides the Math
  + Software engineering creates deployment and service endpoints
  + Cross training is helpful
  + Expect to deal with structured data (e.g., database data, spreadsheet), semi-structured data (e.g., XML, JSON) and or unstructured data (log files, audio, video, social media)
  + The Following well-known Math concepts are frequently used: Linear Algebra; Matrix algebra and eigenvalues; Calculus for Data Science; Derivatives and gradients; Gradient Descent, and a simple neural network
  + The following are some well-known Math packages: Pandas - for data wrangling, NumPy - for large, multi-dimensional arrays, matrices, and high-level mathematical functions, SciPy - for linear algebra, interpolation, optimization, integration, and statistics, Matplotlib- for an object-oriented API for embedding plots into applications, Seaborn for data visualization , Scikit Learn for supervised and unsupervised learning algorithms, and TensorFlow - to create large-scale neural networks
  + To create a model - Adopt the 5E’s model approach - Engage, Explore, Explain, Elaborate, and evaluate
    - Use data access
    - Use regression and classification to make predictions
    - Create Prototype‐based models
    - Apply model to data and validate
    - You start with the internal data - cleanse, wrangle and enrich data
    - Compare old model to new model
    - Compare different models
    - Prevent model bias - use a representative dataset, choose the right model, perform real-world testing, on real-world datasets; Use multiple people to code the data, have participants review results, verify with more data sources, check for alternative explanations, review findings with peers
    - Deploy - AWS cloud, API Gateway control, Direct Connect, Django, Flask, Azure Runbooks, ML Studio
  + Implement Machine Learning as a service to provide real-time prediction.
  + Implement Stream analytics and machine learning capability to provide recommendations and detect fraud/issues

Lessons -

* Business owners and product owners are critical
* Good Data is very important - data cleansing and enrichment are critical
* Knowledge of how to create and use the right model is essential
* Leverage automatic model tuning using tools such as, Autopilker, AWS SageMaker automatic model tuning, Azure HyperParameters
* Data engineers to need to ensure data pipeline with appropriate roadmaps

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

Analyze data quality

Where is data (internal/external)?

Ensure adequate infrastructure is available

Ensure Skills, training, support. Experts can be imported or re-skilled existing personnel

Lean governance. Start with DACI (driver = runs COE, Approver=approves decisions, Contributors = experts, Informed = need to know)

AI team responsibilities – Create vision for ML/AI in the company

External partners such as, universities, vendors and start-ups can be leveraged

Most important roles are – Data Scientist and Data Analyst

Companies such as, DataRobot, universities like MIT and Stanford are offering short executive programs

Organization define structure

Focus on small successes at the outset. Start agile, short-term projects to build confidence and competence

Dedicated resources

Identify use cases

Start governance and compliance at the outset

Be mindful of ethics, bias and legal concerns

The more raw data, the better

Learn when and how to trust your model

Focus on machine learning, instead of big data

COE must be receptive to modifications such as, upgrade in technology, resources, competencies and ML/AI offerings