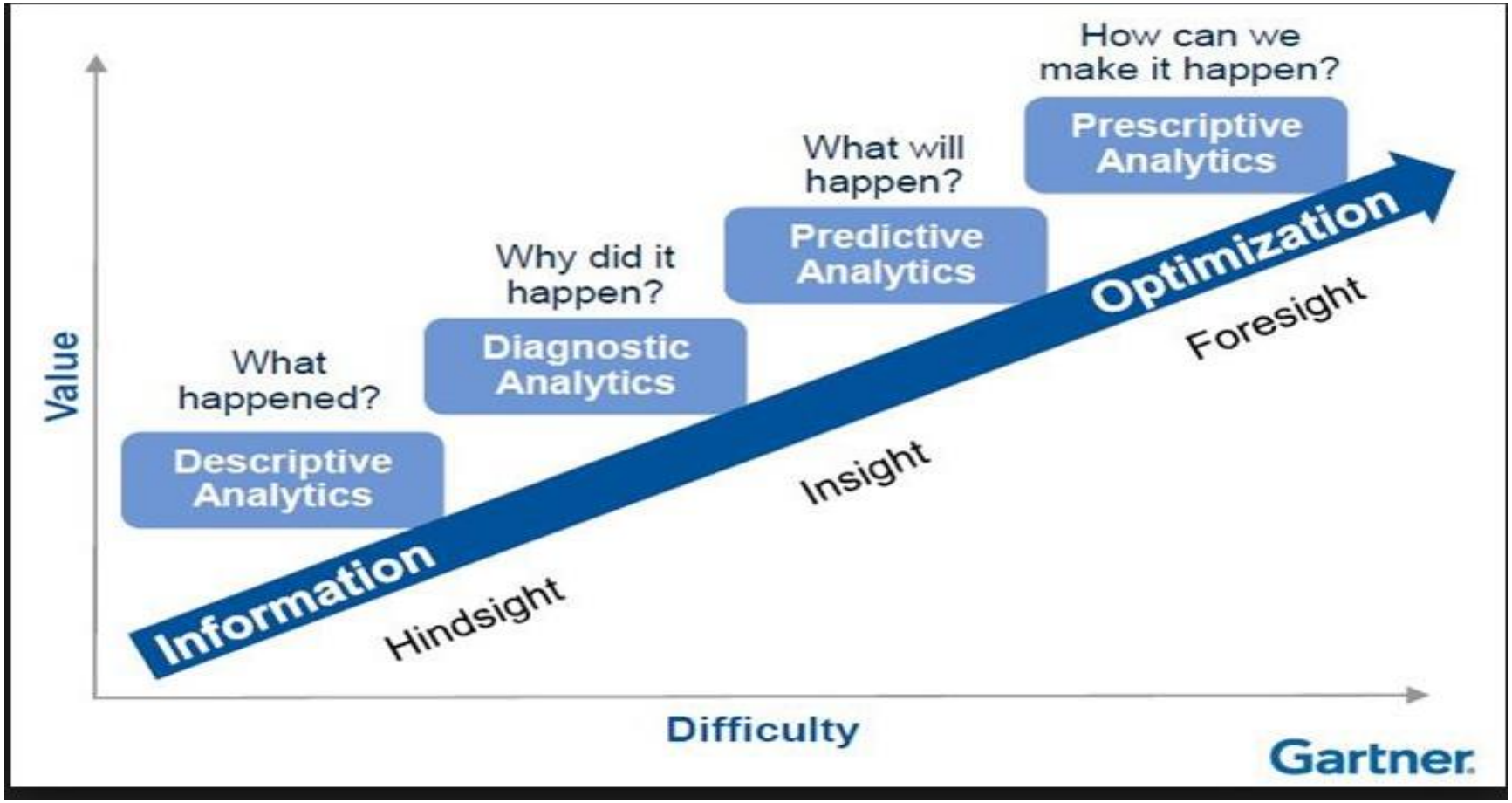
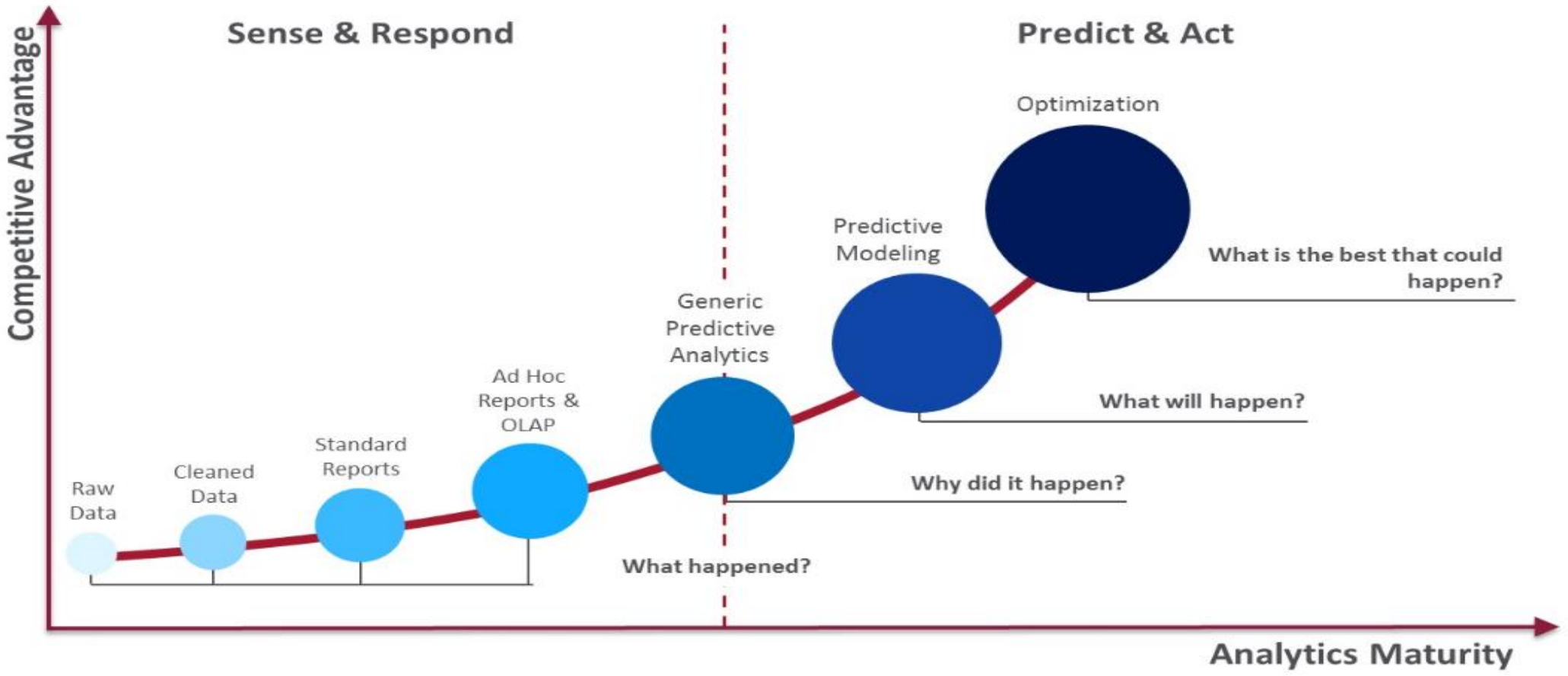
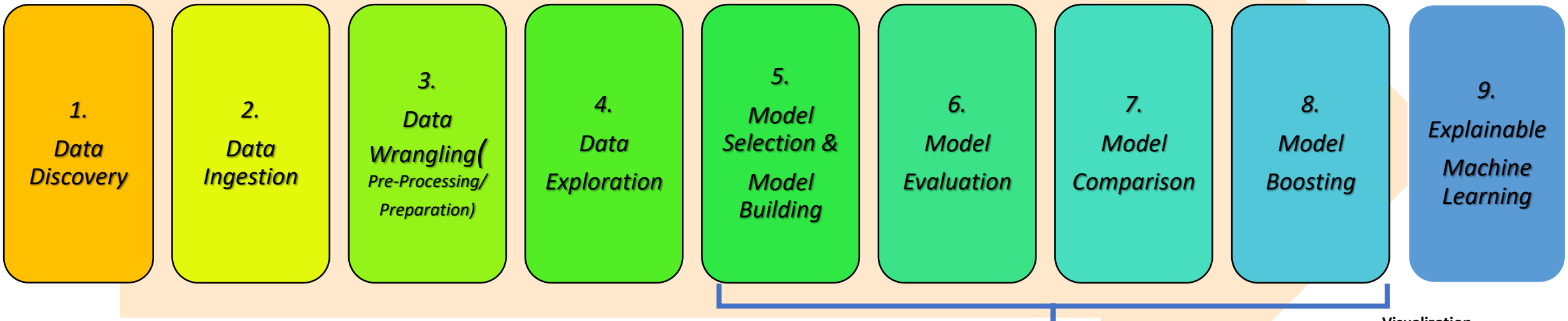
Analytics Maturity in Organizations



Analytics Maturity in Organizations



Data Science Project Life Cycle



- Python
- Pandas
- Numpy
- SQL

- Define Problem statement Requirements
- Assess Data Sources
- Identify Key Business fields for Sampling

- Python
- Kafka
- NiFi
- Spark

- Assess Data Sources
- Batch Extract thru ETL (Extraction Transformation Load)
- Real Time Extract thru Kafka/Ni Fi Spark

- Python
- Pandas
- Numpy
- Spark

Data Cleansing
- Remove Missing Values
- Remove Outliers
- Data Imputation

Data Manipulation
- Rename Columns
- Data Summarization
- Data Fitering Sorting/Grouping
- Merge/Join/Concat

Data Analysis
- Python
- Pandas
- Numpy

Data Visualization
- Matplotlib
- Seaborn

Univariate Analysis
- .info()
- .describe()
- Barplots,histogram,
- Count plots

Bi Variate Analysis
- Scatter Plot
- .Corr
- Correlation Plot
- Regression Plot

- Python
- Pandas
- Numpy

Train/Test Split
- Sci-Kit Learn

Model Fitting
Sci-Kit Learn

Model Prediction
Sci-Kit Learn

Model Visualization
- Matplotlib
- Seaborn

Model Selection/Fitting - Fit model on Training Data
- Derive Correlation / Derive Independent & Dependent variable
- Prepare Train and Test Data
- Decide on Supervised/UnSupervised learning
- Decide on Classification/Regression for supervised
- Decide on Clustering or dimension reduction(pca) for unsupervised
- Fit the data for Multiple model techniques
- If classification – Logistic Reg/Random Forest/Decision Tree/KNN/NB/SVM etc
- If regression – Linear Regression/Random Forest/Decision Tree/KNN..etc

Model Prediction - Predict on Test Data

Model Evaluation - Measures performance of Model

Model Comparison - Compares performance of different model techniques

Model Boosting - Boost the performance of chosen model

Move to Prod – Move the model to production as a .pkl file

- Python
- Pandas
- Numpy

Model Metrics
- Sci-Kit Learn

Model Visualization
- Matplotlib
- Seaborn

- Python
- Pandas
- Numpy

Model Visualization
- Matplotlib
- Seaborn

- Python
- Pandas
- Numpy
- Sci-Kit Learn

Model Visualization
- Matplotlib
- Seaborn

- Python
- Pandas
- Numpy
- Sci-Kit Learn

Model Visualization
- Matplotlib
- Seaborn

Visualization
- Actual Vs Predicted
- Write up on corrective and preventive actions
- Dashboards /Stories on current performance
- Evidence on data anomalies and way to correct .