DATA ANALYTICS

OVERVIEW

Cesar Acosta

Department of Industrial and Systems Engineering

University of Southern California

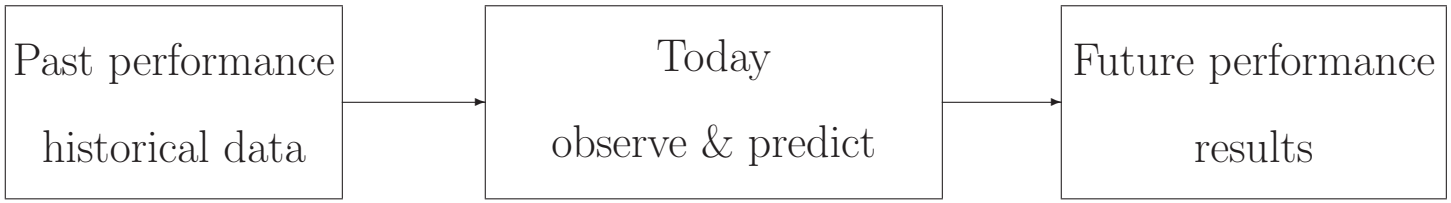
August 24, 2018

DEFINITION

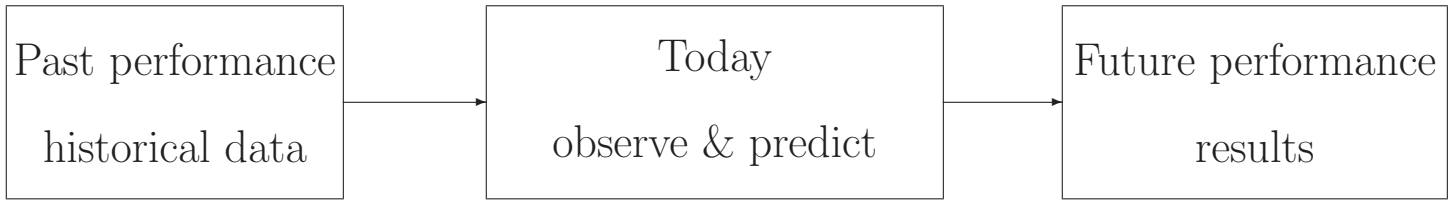
DATA ANALYTICS

Set of mathematical and statistical methods
to collect *meaningful* information
from data

INTRODUCTION



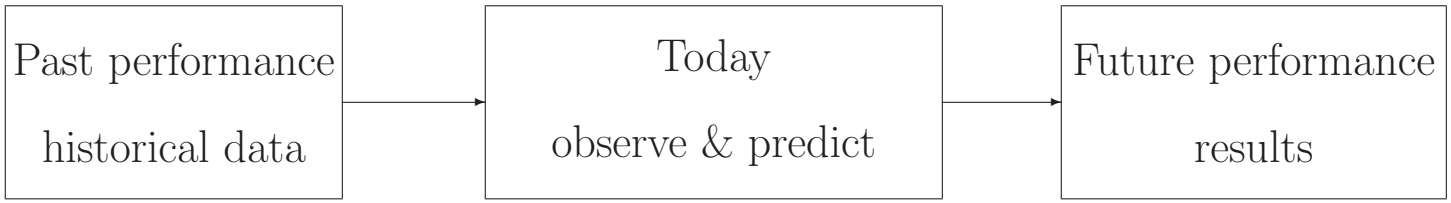
INTRODUCTION



What has happened?

What may happen?

INTRODUCTION



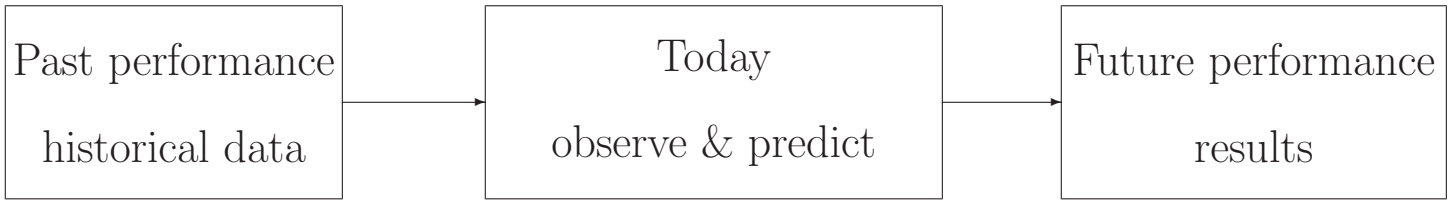
What has happened?

describe/summarize data

What may happen?

scenarios

INTRODUCTION



What has happened?

describe/summarize data

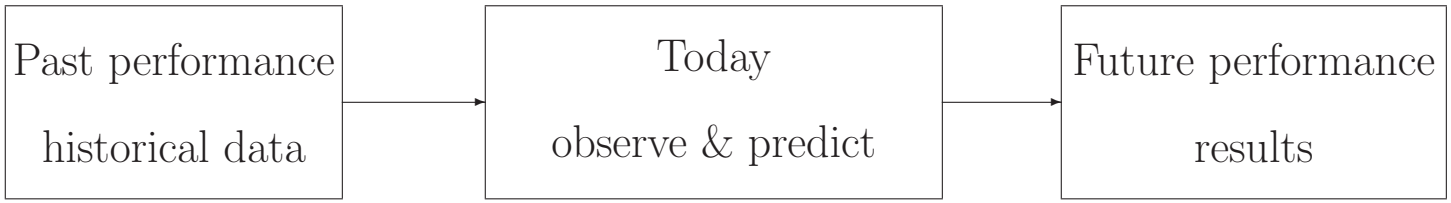
Descriptive Statistics

What may happen?

scenarios

Prediction Models

INTRODUCTION



What has happened?

describe/summarize data

Descriptive Statistics

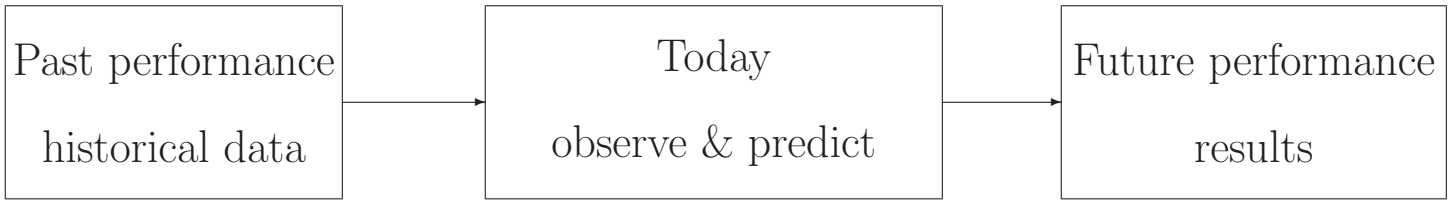
barplots, scatterplots, boxplots
piecharts, line charts
histograms
averages
std. deviations
correlations
histograms

What may happen?

scenarios

Prediction Models

INTRODUCTION



What has happened?

describe/summarize data

Descriptive Statistics

barplots, scatterplots, boxplots
piecharts, line charts
histograms
averages
std. deviations
correlations
histograms

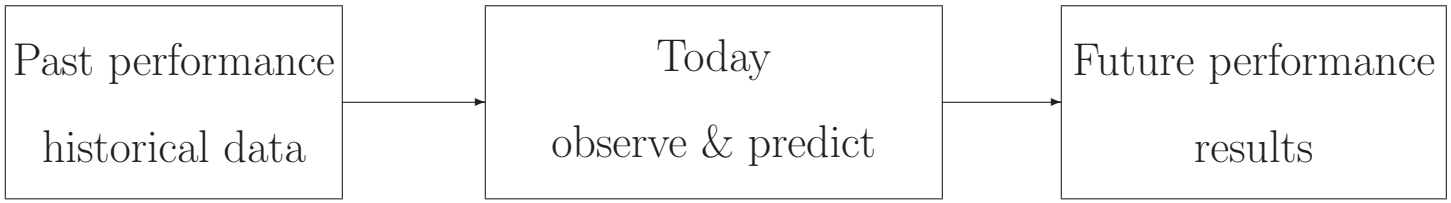
What may happen?

scenarios

Prediction Models

Classification models
Prediction models
Clustering methods

INTRODUCTION



What has happened?

describe/summarize data

Descriptive Statistics

barplots, scatterplots, boxplots
piecharts, line charts
histograms
averages
std. deviations
correlations
histograms

Descriptive Analytics

What may happen?

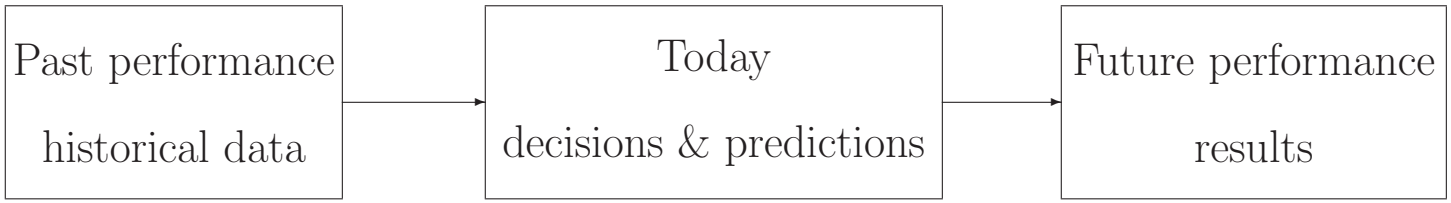
scenarios

Prediction Models

Classification models
Prediction models
Clustering methods

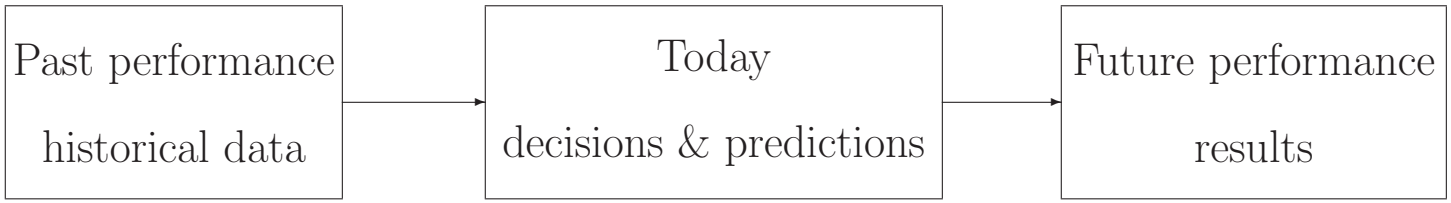
Predictive Analytics

INTRODUCTION



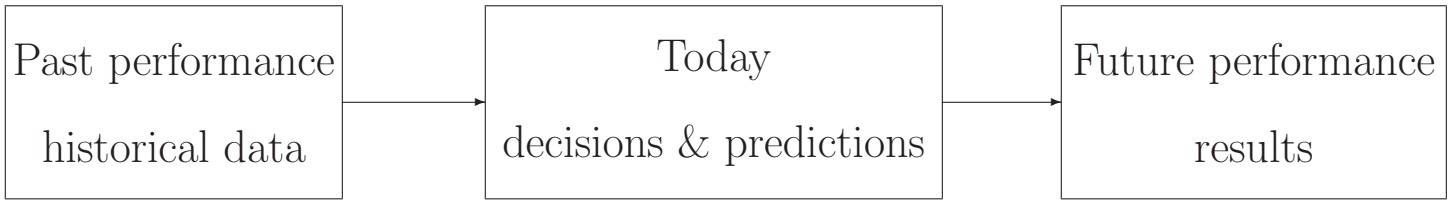
What has happened?	What we want to happen?	What may happen?
describe/summarize data		scenarios
Descriptive Statistics		Prediction Models
barplots, scatterplots, boxplots piecharts, line charts histograms averages std. deviations correlations histograms		Classification models Prediction models Clustering methods
Descriptive Analytics		Predictive Analytics

INTRODUCTION



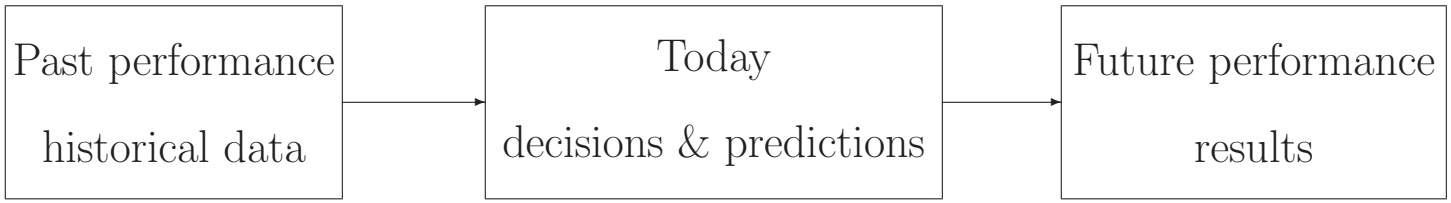
What has happened?	What we want to happen?	What may happen?
describe/summarize data	What decisions	scenarios
Descriptive Statistics	are needed today	Prediction Models
barplots, scatterplots, boxplots piecharts, line charts histograms averages std. deviations correlations histograms	to make things happen?	Classification models Prediction models Clustering methods
Descriptive Analytics		Predictive Analytics

INTRODUCTION



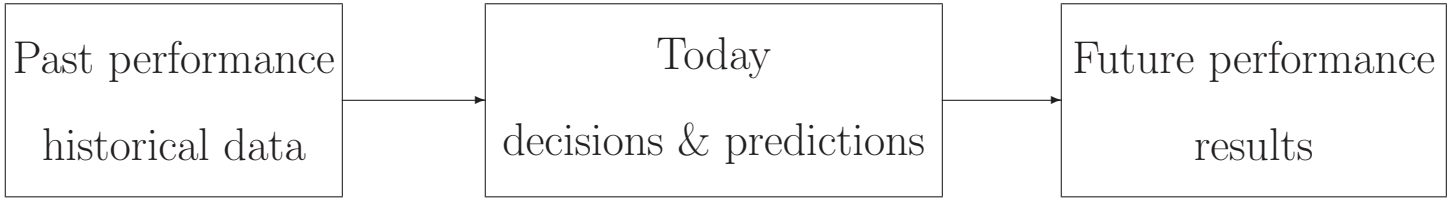
What has happened?	What we want to happen?	What may happen?
describe/summarize data	What decisions	scenarios
Descriptive Statistics	are needed today	Prediction Models
barplots, scatterplots, boxplots piecharts, line charts histograms averages std. deviations correlations histograms	to make things happen? Simulation models Optimization models	Classification models Prediction models Clustering methods
Descriptive Analytics		Predictive Analytics

INTRODUCTION



What has happened?	What we want to happen?	What may happen?
describe/summarize data	What decisions	scenarios
Descriptive Statistics	are needed today	Prediction Models
barplots, scatterplots, boxplots piecharts, line charts histograms averages std. deviations correlations histograms	to make things happen? Simulation models Optimization models	Classification models Prediction models Clustering methods
Descriptive Analytics	Prescriptive Analytics	Predictive Analytics

INTRODUCTION



What has happened?	What we want to happen?	What may happen?
describe/summarize data	What decisions	scenarios
Descriptive Statistics	are needed today	Prediction Models
barplots, scatterplots, boxplots piecharts, line charts histograms averages std. deviations correlations histograms	to make things happen? Simulation models Optimization models	Classification models Prediction models Clustering methods
Descriptive Analytics	Prescriptive Analytics	Predictive Analytics

DESCRIPTIVE ANALYTICS

- Summary of historical data

DESCRIPTIVE ANALYTICS

- Summary of historical data

- We need
  - clean
  - up to date
  - well organized

  data sets

DESCRIPTIVE ANALYTICS

- Summary of historical data

- We need
  - clean
  - up to date
  - well organized

  data sets

Business Intelligence software

DESCRIPTIVE ANALYTICS - BUSINESS INTELLIGENCE (BI)

- Software
  - to collect data from internal systems
  - to prepare it for descriptive analysis
  - to create reports and dashboards
  - in very efficient ways

DESCRIPTIVE ANALYTICS - BI SOFTWARE

- Tableau
- MS Power BI
- Qlik
- SAS
- SAP
- IBM Cognos

GARTNER -BI ANALYTICS- MAGIC QUADRANT

DESCRIPTIVE ANALYTICS - BI SOFTWARE

- Tableau
- MS Power BI
- Qlik
- SAS
- SAP
- IBM Cognos

Very limited for Predictive Analytics

PREDICTIVE ANALYTICS - MODELING

- PREDICTION
- CLASSIFICATION
- CLUSTERING

PREDICTIVE ANALYTICS - MODELING

- PREDICTION
- CLASSIFICATION
- CLUSTERING

Machine learning software

PREDICTIVE ANALYTICS - MACHINE LEARNING SOFTWARE

- SAS, R, Weka
- RHadoop, SAS over Hadoop
- Spark, Twister, Haloop, GraphLab
- SAS in-memory Analytics

PREDICTIVE ANALYTICS - MACHINE LEARNING SOFTWARE

- SAS, R, Weka (Most powerful)
- RHadoop, SAS over Hadoop (Big data)
- Spark, Twister, Haloop, GraphLab (Big data, 2nd generation)
- SAS in-memory Analytics (Big data, 3rd generation)

PREDICTIVE ANALYTICS - PREDICTION MODELS

- MLR with numerical/categorical predictors
- Nonlinear regression
  - Polynomial regression
  - Interaction terms
- Generalized Linear Models
- Regression Trees
- Support Vector Machine for regression

PREDICTIVE ANALYTICS - CLASSIFICATION MODELS

- Logistic regression
- KNN
- Classification Trees,
- Ensemble of Trees (Bagging, Boosting, Random Forest)
- SVM

PREDICTIVE ANALYTICS - CLUSTERING METHODS

- Principal Components
- K-means
- Hierarchical clustering

DATA ANALYTICS - Machine Learning

- Graph Search
- Voice Recognition
- Recommender System
- Matching
- Text classification (fake?)