

DATA ANALYTICS

OVERVIEW

Cesar Acosta

Department of Industrial and Systems Engineering

University of Southern California

August 25, 2017

DEFINITION

DATA ANALYTICS

Set of mathematical and statistical methods
to generate meaningful information
from data sets

DEFINITION

DATA ANALYTICS

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

DEFINITION

DATA ANALYTICS

meaningful information
useful for decision making

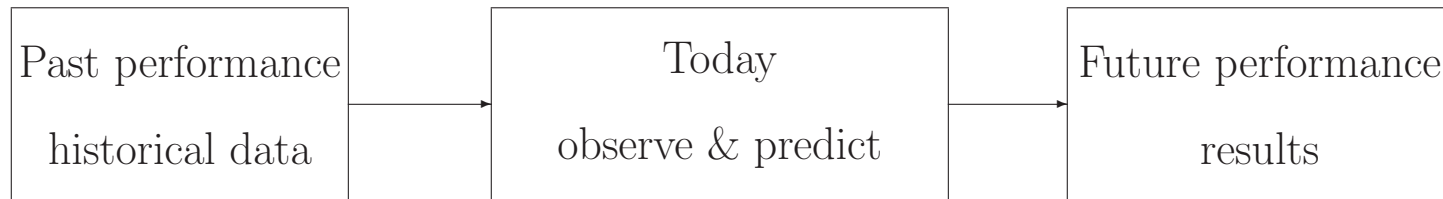
DEFINITION

DATA ANALYTICS

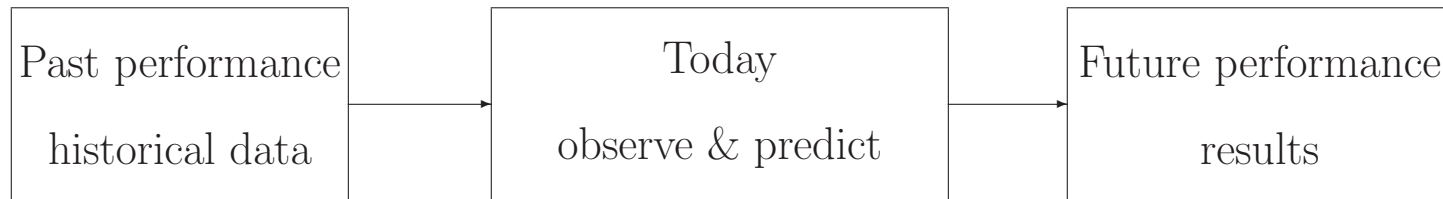
meaningful information
useful for decision making

- Basic statistics (one population)
- Trend (one population)
- Comparisons/Correlations (2+ populations)
- Patterns
- Outliers
- Associations
- Predictions

INTRODUCTION



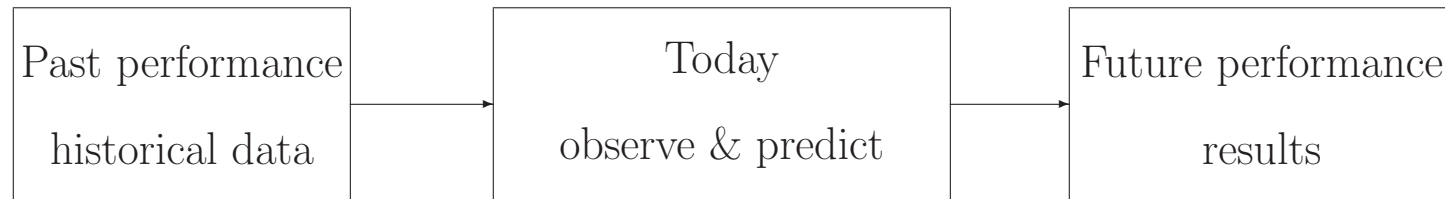
INTRODUCTION



What has happened?

What may happen?

INTRODUCTION



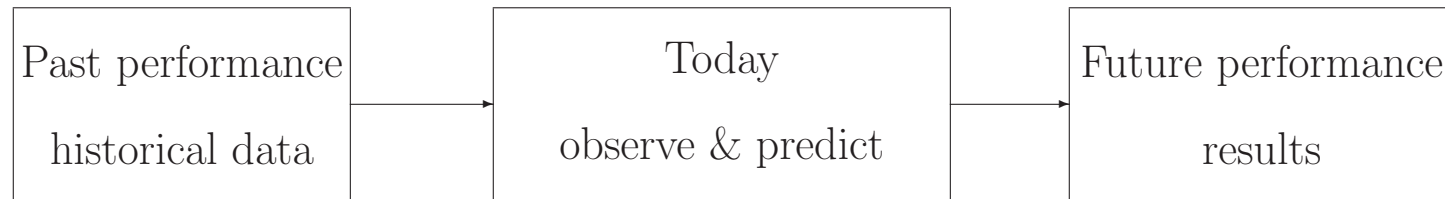
What has happened?

describe/summarize data

What may happen?

scenarios

INTRODUCTION



What has happened?

describe/summarize data

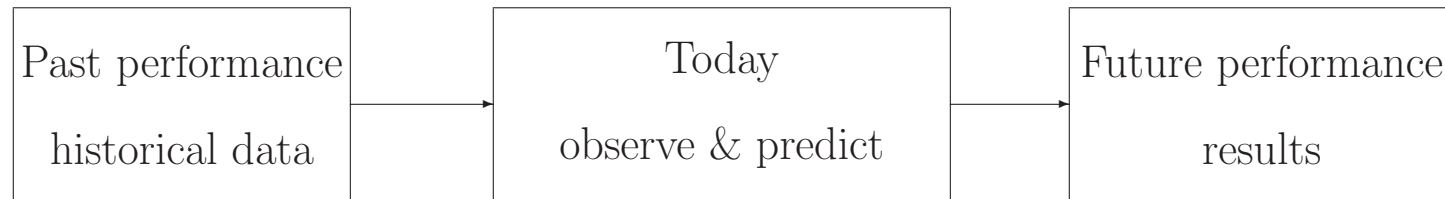
Descriptive Statistics

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

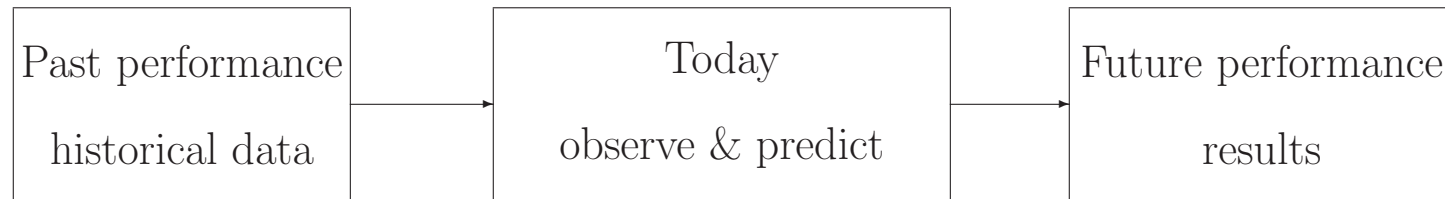
histograms

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

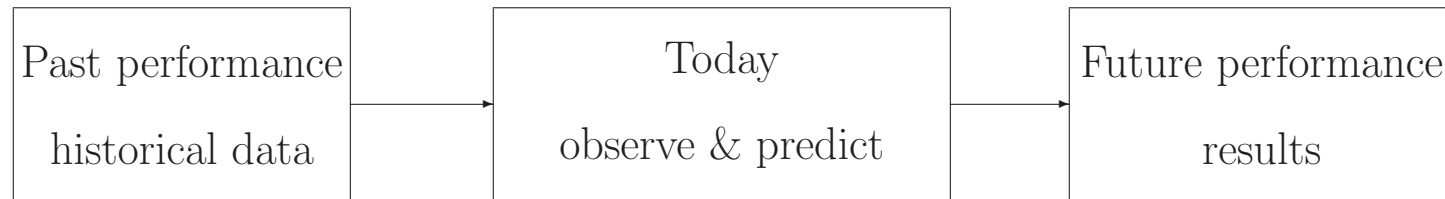
Prediction Models

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

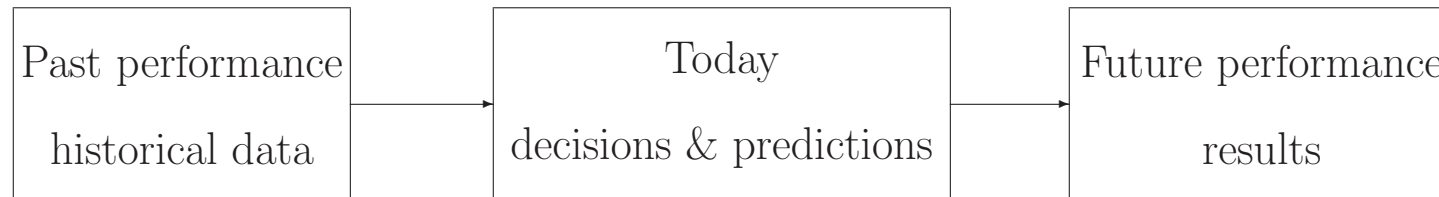
Classification models

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

Classification models

piecharts, line charts

Prediction models

histograms

Clustering methods

averages

std. deviations

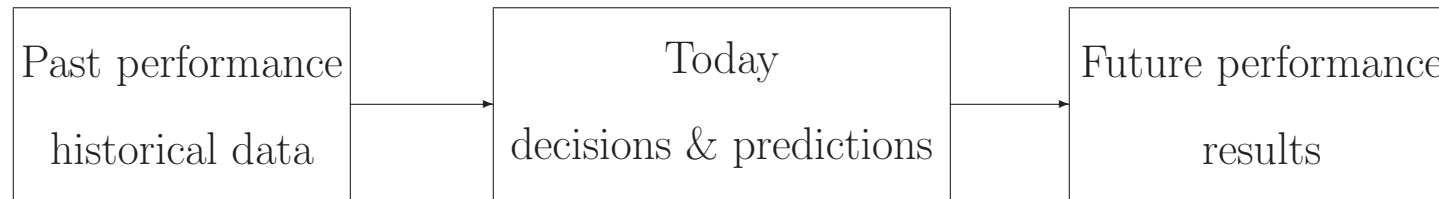
correlations

histograms

Descriptive Analytics

Predictive Analytics

INTRODUCTION



What has happened?

describe/summarize data

Descriptive Statistics

barplots, scatterplots, boxplots

piecharts, line charts

histograms

averages

std. deviations

correlations

histograms

Descriptive Analytics

What we want to happen?

What decisions

are needed today

to make things happen?

What may happen?

scenarios

Prediction Models

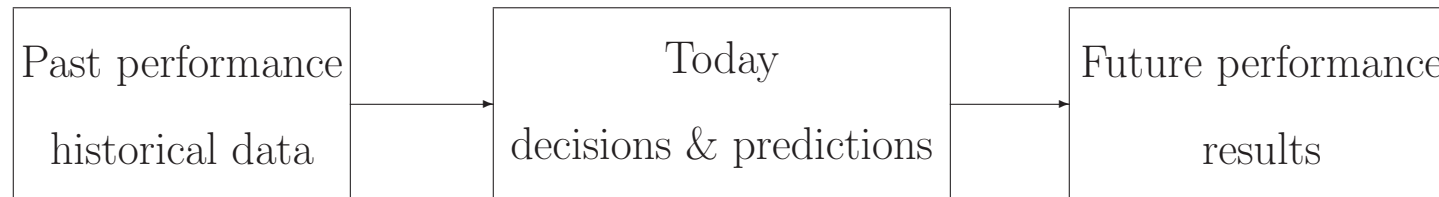
Classification models

Prediction models

Clustering methods

Predictive Analytics

INTRODUCTION



What has happened?

describe/summarize data

Descriptive Statistics

barplots, scatterplots, boxplots

piecharts, line charts

histograms

averages

std. deviations

correlations

histograms

Descriptive Analytics

What we want to happen?

What decisions

are needed today

to make things happen?

Simulation models

Optimization models

What may happen?

scenarios

Prediction Models

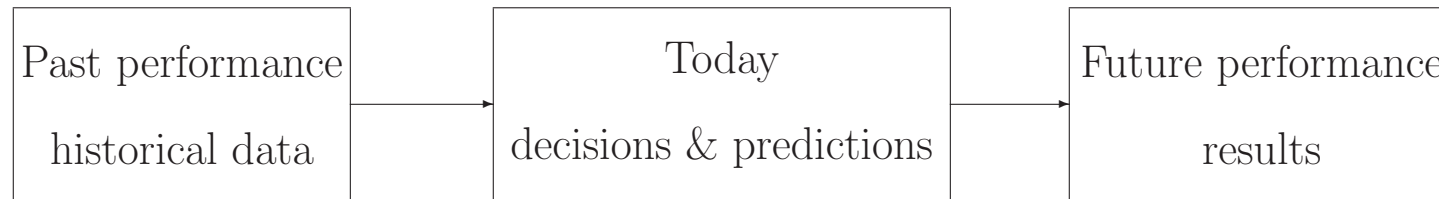
Classification models

Prediction models

Clustering methods

Predictive Analytics

INTRODUCTION



What has happened?

describe/summarize data

Descriptive Statistics

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

Descriptive Analytics

What we want to happen?

What decisions

are needed today

to make things happen?

Simulation models

Optimization models

Prescriptive Analytics

What may happen?

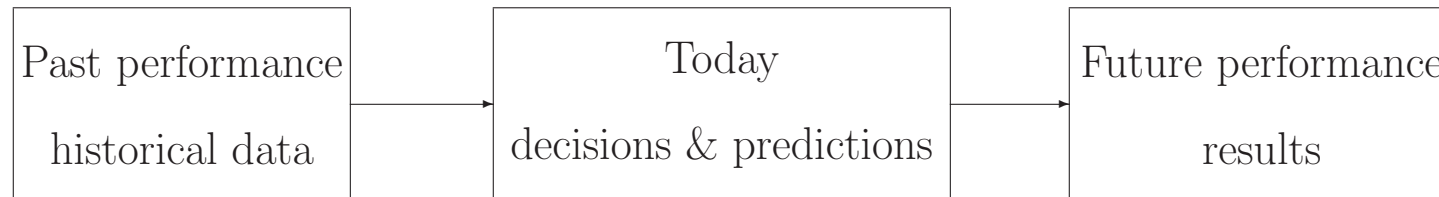
scenarios

Prediction Models

Classification models
Prediction models
Clustering methods

Predictive Analytics

INTRODUCTION



What has happened?

describe/summarize data

Descriptive Statistics

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

Descriptive Analytics

What we want to happen?

What decisions

are needed today

to make things happen?

Simulation models

Optimization models

Prescriptive Analytics

What may happen?

scenarios

Prediction Models

Classification models
Prediction models
Clustering methods

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

GARTNER -BI ANALYTICS- MAGIC QUADRANT



DESCRIPTIVE ANALYTICS - BI SOFTWARE

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

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, Hadoop, GraphLab
- SAS in-memory Analytics

PREDICTIVE ANALYTICS - MACHINE LEARNING SOFTWARE

- SAS, R, Weka (Most powerful)
- RHadoop, SAS over Hadoop (Big data)
- Spark, Twister, Hadoop, 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
- Nonparametric regression
- Regression Trees
- SVM regression

PREDICTIVE ANALYTICS - CLASSIFICATION MODELS

- Logistic regression
- KNN
- Classification Trees, Bagging, RF, Boosting
- SVM

PREDICTIVE ANALYTICS - CLUSTERING METHODS

- Principal Components
- K-means
- Hierarchical clustering