



# **SKILLS BOOTCAMP IN DIGITAL: DATA ENGINEERING AND ANALYTICAL TOOLS.**

**Industry Standard Tools:  
Analytics Models.**

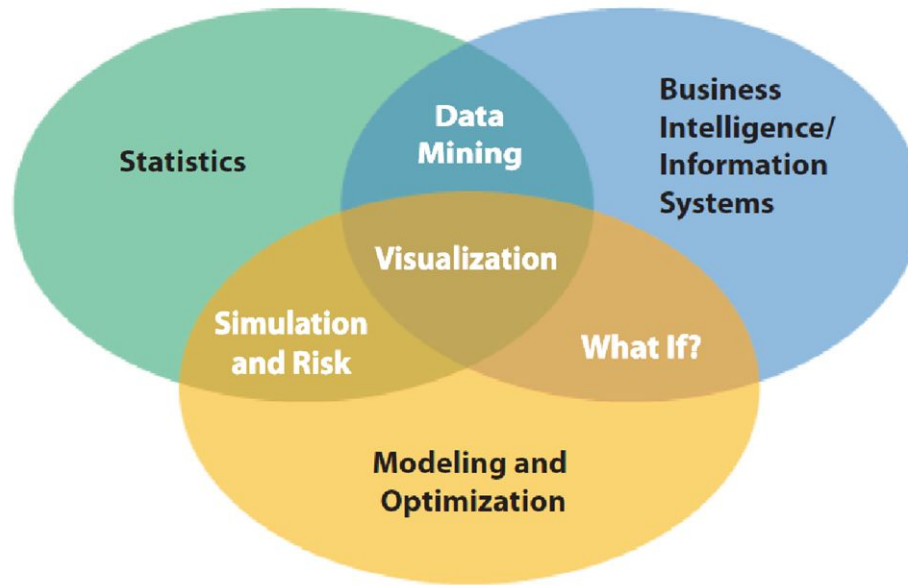
**[monika.sharma@gateshead.ac.uk](mailto:monika.sharma@gateshead.ac.uk)**

# Learning Outcomes

- Explain the difference between descriptive, predictive and prescriptive analytics
- Explain the concept of a model and the various ways a model can be characterised
- Understand how to implement models in R



# A Visual Perspective of Business Analytics



# Descriptive, Predictive, and Prescriptive Analytics

- Descriptive analytics:
  - the use of data to understand past and current business performance and make informed decisions
- Predictive analytics:
  - predict the future by examining historical data, detecting patterns or relationships in these data, and then extrapolating these relationships forward in time.
- Prescriptive analytics:
  - identify the best alternatives to minimize or maximize some objective



# Big Data

- Big data refers to massive amounts of business data (volume) from a wide variety of sources (variety), much of which is available in real time (velocity), and much of which is uncertain or unpredictable (veracity).
- “The effective use of big data has the potential to transform economies, delivering a new wave of productivity growth and consumer surplus. Using big data will become a key basis of competition for existing companies, and will create new competitors who are able to attract employees that have the critical skills for a big data world.”

- McKinsey Global Institute, 2011



# Data Reliability and Validity

- Reliability - data is accurate and consistent.
- Validity - data correctly measures what it is supposed to measure.
- Examples:
  - A tyre pressure gage that consistently reads several pounds of pressure below the true value is not reliable, although it is valid because it does measure tyre pressure
  - The number of calls to a customer service desk might be counted correctly each day (and thus is a reliable measure) but not valid if it is used to assess customer dissatisfaction, as many calls may be simple queries.
  - A survey question that asks a customer to rate the quality of the food in a restaurant may be neither reliable (because different customers may have conflicting perceptions) nor valid (if the intent is to measure customer satisfaction, as satisfaction generally includes other elements of service besides food).



# Models in Analytics

- Model - an abstraction or representation of a real system, idea, or object.
  - Captures the most important features
  - Can be a written or verbal description, a visual representation, a mathematical formula, or a spreadsheet.



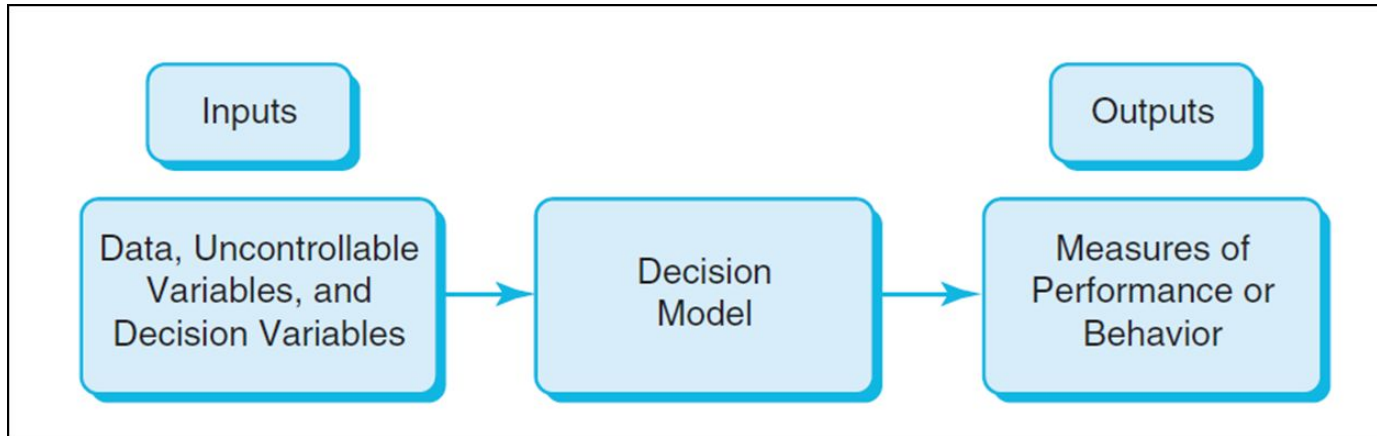
# Decision Models

- Decision Model - a logical or mathematical representation of a problem or business situation that can be used to understand, analyze, or facilitate making a decision
- Inputs:
  - Data - assumed to be constant
  - Uncontrollable inputs - quantities that can change but cannot be controlled
  - Decision options - controllable and selected at the discretion of the decision maker





# Nature of Decision Models

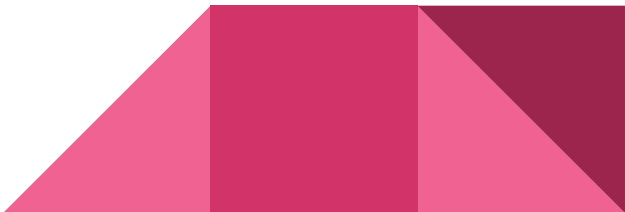


# Descriptive Models

- **Descriptive models** explain behavior and/or seeks to describe what is happening



# How could we use R for descriptive analytics?

- Minimum and maximum
  - Range
  - Mean
  - Median
  - First and third quartile
  - Other quantiles
  - Interquartile range
  - Standard deviation and variance
  - Summary
  - Barplot
  - Histogram
  - Boxplot
  - Dotplot
  - Scatterplot
  - Line plot
  - QQ-plot
  - Density plot
  - Correlation plot
- 

# Predictive Models

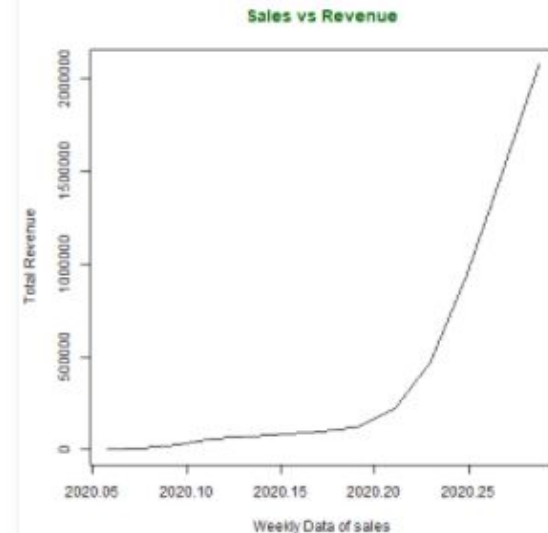
- **Predictive models** focus on what will happen in the future.
- Many predictive models are developed by analyzing historical data and assuming that the past is representative of the future.



# How could we use R for predictive analytics?

## Time Series Analysis

```
x <- c(580, 7813, 28266, 59287, 75700,  
      87820, 95314, 126214, 218843, 471497,  
      936851, 1508725, 2072113)  
  
# library required for decimal_date() function  
library(lubridate)  
  
# output to be created as png file  
png(file = "predictiveAnalysis.png")  
  
# creating time series object  
# from date 22 January, 2020  
mts <- ts(x, start = decimal_date(ymd("2020-01-22")),  
          frequency = 365.25 / 7)  
  
# plotting the graph  
plot(mts, xlab = "Weekly Data of sales",  
      ylab = "Total Revenue",  
      main = "Sales vs Revenue",  
      col.main = "darkgreen")  
  
# saving the file  
dev.off()
```



from GeeksForGeeks

# How could we use R for predictive analytics?

## Forecasting

```
x <- c(580, 7813, 28266, 59287, 75700,
      87820, 95314, 126214, 218843,
      471497, 936851, 1508725, 2072113)

# library required for decimal_date() function
library(lubridate)

# library required for forecasting
library(forecast)

# output to be created as png file
png(file = "forecastSalesRevenue.png")

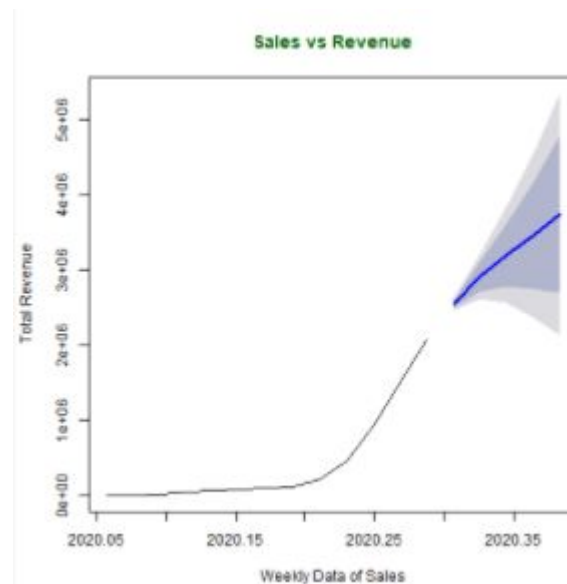
# creating time series object
# from date 22 January, 2020
mts <- ts(x, start = decimal_date(ymd("2020-01-22")),
          frequency = 365.25 / 7)

# forecasting model using arima model
fit <- auto.arima(mts)

# Next 5 forecasted values
forecast(fit, 5)

# plotting the graph with next
# 5 weekly forecasted values
plot(forecast(fit, 5), xlab = "Weekly Data of Sales",
     ylab = "Total Revenue",
     main = "Sales vs Revenue", col.main = "darkgreen")

# saving the file
dev.off()
```



# Prescriptive Models

- Prescriptive models help decision makers identify the best solution to a decision problem.
- Optimization - finding values of decision variables that minimize (or maximize) something such as cost (or profit)
  - Objective function - the equation that minimizes (or maximizes) the quantity of interest
  - Optimal solution - values of the decision variables at the minimum (or maximum) point



# Activities

- Attempt the examples of descriptive and predictive analytics mentioned on slides 12, 14 +15
- Complete any outstanding R Primers on R Studio Cloud

