

## Contents

<i>Preface</i>	(ix)
<i>About the Author</i>	(xi)

---

### Chapter 1: Data Science: Key Concepts

<i>Five Disruptive Products</i> .....	1
<i>The Japanese App</i> .....	1
<i>The Healthcare App</i> .....	2
<i>Coursera</i> .....	3
<i>Sensory Device in Agriculture Sector</i> .....	4
<i>Autonomous Car</i> .....	4
<i>Data Science Vs Traditional Methods</i> .....	5
<i>Differences in Architecture</i> .....	6
<i>Demystifying Machine Learning</i> .....	7
<i>Technique 1: Segmentation</i> .....	7
<i>Technique 2: Unstructured Text mining</i> .....	11
<i>Technique 3: Use of scoring for signal processing</i> .....	12
<i>Technique 4: Forecasting</i> .....	12
<i>Technique 5: Recommenders</i> .....	12
<i>Reference Architecture</i> .....	14
<i>Hands on Segementation</i> .....	16
<i>Summary</i> .....	18

---

### Chapter 2: Spotting Signals: An Overview

<i>Signals across Verticals</i> .....	20
<i>Analyzing a Signal Pattern</i> .....	21
<i>Application of Signal Pattern Observed</i> .....	21
<i>Signals: A Few Key Concepts</i> .....	23
<i>Signal Extraction Methodology – Simplistic View</i> .....	24
<i>Simplistic Nine Step Process</i> .....	24
<i>Analytical Tool – Tool R</i> .....	26
<i>Basic Commands in R</i> .....	27
<i>Step 1 Download R</i> .....	27

---

### Chapter 3: Problem based Analysis

<i>Exploratory Data Analysis (EDA)</i> .....	35
<i>Univariate Analysis</i> .....	35
<i>Questions Answered by Univariate Analysis</i> .....	36
<i>Objectives of EDA</i> .....	37
<i>Univariate Analysis Across Sectors</i> .....	38
<i>Types of Variables and Measures in EDA</i> .....	39
<i>Univariate Data Analysis</i> .....	40
<i>Measures of Central Tendency</i> .....	40
<i>Measure of Dispersion</i> .....	40
<i>Scope of Data Analyzed - EDA (Fleet Industry)</i> .....	42
<i>Basic Commands in R</i> .....	47
<i>Summary</i> .....	49

---

### Chapter 4: Bivariate Analysis

<i>Correlation</i> .....	51
<i>Cross Tabs</i> .....	52
<i>Cross Tab Analysis – Core Philosophy</i> .....	55
<i>Correlation Analysis Across Industries</i> .....	55
<i>Retail Industry</i> .....	55
<i>Telecom Industry</i> .....	56
<i>Banking Industry</i> .....	57
<i>Cross Tab Analysis Across Industries</i> .....	57
<i>Retail Industry</i> .....	57
<i>Telecom Industry</i> .....	57
<i>Banking Industry</i> .....	58
<i>Pearson's Correlation</i> .....	58
<i>Core Foundational Concepts in Cross Tabs</i> .....	59
<i>Summary</i> .....	62

---

### Chapter 5: Visual Constructs

<i>Detecting Patterns using Visual Constructs</i> .....	63
<i>Demystifying Advanced Visualization</i> .....	64
<i>Box Plot</i> .....	64
<i>Scatter Plot</i> .....	65
<i>Geospatial Map</i> .....	65
<i>Heat Maps</i> .....	66

<i>Spider Chart</i> .....	68
<i>Advanced Visualization Example</i> .....	68
<i>Domestic Loan Analysis</i> .....	69
<i>Core Concepts in Advanced Visualization</i> .....	70
<i>Visualization Consumers</i> .....	70
<i>Creating Dashboards</i> .....	71
<i>Connecting the Dots to Create a Dashboard</i> .....	72
<i>Best Practices in Designing Dashboards and Scoreboards</i> .....	72
<i>Visualization Commands in R</i> .....	73
<i>Summary</i> .....	76

---

## **Chapter 6: Business Story Telling using R**

<i>Visualization Commands in R</i> .....	77
<i>Small Multiples: A Visualization Construct</i> .....	79
<i>Bubble Chart: A Visualization Construct</i> .....	81
<i>10 Basic Commands in R</i> .....	82
<i>Summary</i> .....	85

---

## **Chapter 7: Exploratory Data Analysis Case Study**

<i>Exploratory Data Analysis</i> .....	89
<i>Scenario 1: Survival Ananalysis</i> .....	89
<i>Scenario 2: Attrition Ananalysis</i> .....	91
<i>Scenario 3: Difference between Active and Inactive Customers: Valuable Vulnerable</i> .....	94
<i>Scenario 4: Days to Repeat Purchases</i> .....	96
<i>Scenario 5: Sales Trends. Seasonality in Sales/ Identifying Patterns</i> .....	99
<i>Scenario 6: Segmenting Watch Company's Customers Region Wise</i> .....	100
<i>Scenario 7: Customer Lifetime Value</i> .....	103
<i>Summary</i> .....	104

---

## **Chapter 8: Machine Learning in Action**

<i>Support Vector Machines (SVM)</i> .....	105
<i>Decision Tree</i> .....	106
<i>Three Important Things about Decision</i> .....	107

<i>Random Forest</i> .....	107
<i>A/B Testing</i> .....	108
<i>Collaborative Filtering</i> .....	110
<i>Fixed Size Neighborhood</i> .....	111
<i>Threshold based Neighbourhood</i> .....	112
<i>Graph</i> .....	114
<i>Structure Unstructured Data</i> .....	117
<i>Summary</i> .....	118

---

## **Chapter 9: Regression**

<i>Introduction</i> .....	119
<i>5 Powerful Unanswered Questions by Regression –</i>	
<i>Unknown Unknowns</i> .....	121
<i>Regression Across Sectors</i> .....	122
<i>Scenario 1: Cost of Insurance</i> .....	122
<i>Scenario 2: Model Building for Property Design</i> .....	123
<i>Scenario 3: Estimating Patients Stay at Hospital</i> .....	124
<i>Scenario 4: Estimate Defect Density</i> .....	124
<i>Population and Sample Regression Models</i> .....	125
<i>Commands in R</i> .....	128
<i>Correlation # Causation</i> .....	130
<i>Summary</i> .....	130

---

## **Chapter 10: Dimensionality Reduction Technique**

<i>Feature Engineering</i> .....	133
<i>Feature Engineering – Key Point</i> .....	135
<i>Feature Selection - Definition</i> .....	135
<i>Feature Selection - Optimality</i> .....	136
<i>Correlation and Variables</i> .....	136
<i>Ranking Criteria - Correlation</i> .....	137
<i>Feature Subset Selection</i> .....	139
<i>Summary</i> .....	142

**Case Studies** ..... 143

**References** ..... 169

**Index** ..... 171