Practices of Business Intelligence

Descriptive Analytics II: Business Intelligence and Data Warehousing

Min-Yuh Day
Assistant Professor
Dept. of Information Management, Tamkang University

http://mail.tku.edu.tw/myday/
2018-10-17
<table>
<thead>
<tr>
<th>週次 (Week)</th>
<th>日期 (Date)</th>
<th>內容 (Subject/Topics)</th>
</tr>
</thead>
</table>
| 1 2018/09/12 | 商業智慧實務課程介紹  
(Course Orientation for Practices of Business Intelligence) |
| 2 2018/09/19 | 商業智慧、分析與資料科學  
(Business Intelligence, Analytics, and Data Science) |
| 3 2018/09/26 | 人工智慧、大數據與雲端運算  
(ABC: AI, Big Data, and Cloud Computing) |
| 4 2018/10/03 | 描述性分析I：數據的性質、統計模型與可視化  
(Descriptive Analytics I: Nature of Data, Statistical Modeling, and Visualization) |
| 5 2018/10/10 | 國慶紀念日 (放假一天)  
(National Day)  
(Day off) |
| 6 2018/10/17 | 描述性分析II：商業智慧與資料倉儲  
(Descriptive Analytics II: Business Intelligence and Data Warehousing) |
課程大綱 (Syllabus)

週次 (Week) 日期 (Date) 內容 (Subject/Topics)
7 2018/10/24 預測性分析 I：資料探勘流程、方法與演算法
   (Predictive Analytics I: Data Mining Process, Methods, and Algorithms)
8 2018/10/31 預測性分析 II：文本、網路與社群媒體分析
   (Predictive Analytics II: Text, Web, and Social Media Analytics)
9 2018/11/07 期中報告 (Midterm Project Report)
10 2018/11/14 期中考試 (Midterm Exam)
11 2018/11/21 處方性分析：最佳化與模擬
   (Prescriptive Analytics: Optimization and Simulation)
12 2018/11/28 社會網絡分析
   (Social Network Analysis)
課程大綱 (Syllabus)

週次 (Week) 日期 (Date) 內容 (Subject/Topics)
13 2018/12/05 機器學習與深度學習
      (Machine Learning and Deep Learning)
14 2018/12/12 自然語言處理
      (Natural Language Processing)
15 2018/12/19 AI交談機器人與對話式商務
      (AI Chatbots and Conversational Commerce)
16 2018/12/26 商業分析的未來趨勢、隱私與管理考量
      (Future Trends, Privacy and Managerial Considerations in Analytics)
17 2019/01/02 期末報告 (Final Project Presentation)
18 2019/01/09 期末考試 (Final Exam)
Business Intelligence (BI)

1. Introduction to BI and Data Science
2. Descriptive Analytics
3. Predictive Analytics
4. Prescriptive Analytics
5. Big Data Analytics
6. Future Trends
Descriptive Analytics II: Business Intelligence and Data Warehousing
Outline

• Descriptive Analytics II
• Business Intelligence
• Data Warehousing
• Data Integration and the Extraction, Transformation, and Load (ETL) Processes
• Business Performance Management (BPM)
• Performance Measurement
  – Balanced Scorecards
  – Six Sigma
Relationship between Business Analytics and BI, and BI and Data Warehousing

Business Analytics

Descriptive

- What happened?
- What is happening?
- Business reporting
- Dashboards
- Scorecards
- Data warehousing
- Well-defined business problems and opportunities

Predictive

- What will happen?
- Why will it happen?
- Data mining
- Text mining
- Web/media mining
- Forecasting
- Accurate projections of future events and outcomes

Prescriptive

- What should I do?
- Why should I do it?
- Optimization
- Simulation
- Decision modeling
- Expert systems
- Best possible business decisions and actions

Business Intelligence

Advanced Analytics

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
A List of Events That Led to Data Warehousing Development

- **1970s**
  - Mainframe computers
  - Simple data entry
  - Routine reporting
  - Primitive database structures
  - Teradata Incorporated

- **1980s**
  - Centralized data storage
  - Data warehousing was born
  - Inmon, *Building the Data Warehouse*
  - Kimball, *The Data Warehouse Toolkit*
  - EDW architecture design

- **1990s**
  - Big Data analytics
  - Social media analytics
  - Text and Web analytics
  - Hadoop, MapReduce, NoSQL
  - In-memory, in-database

- **2000s**
  - Mini/personal computers (PCs)
  - Business applications for PCs
  - Distributer DBMS
  - Relational DBMS
  - Teradata ships commercial DBs
  - *Business Data Warehouse* coined

- **2010s**
  - Exponentially growing Web data
  - Consolidation of DW/BI industry
  - Data warehouse appliances emerged
  - Business intelligence popularized
  - Data mining and predictive modeling
  - Open source software
  - SaaS, PaaS, Cloud computing

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*, 4th Edition, Pearson
Characteristics of Data Warehousing

• Subject oriented
  – Data are organized by detailed subject, such as sales, products, or customers, containing only information relevant for decision support.

• Integrated
  – Integration is closely related to subject orientation.

• Time variant (time series)
  – A warehouse maintains historical data.

• Nonvolatile
  – After data are entered into a data warehouse, users cannot change or update the data.

Data-Driven Decision Making—
Business Benefits of the Data Warehouse

Data Warehouse
One management and analytical platform for product configuration, warranty, and diagnostic readout data

- Reduced Infrastructure Expense
  2/3 cost reduction through data mart consolidation

- Produced Warranty Expenses
  Improved reimbursement accuracy through improved claim data quality

- Improved Cost of Quality
  Faster identification, prioritization, and resolution of quality issues

- Accurate Environmental Performance Reporting

- IT Architecture Standardization
  One strategic platform for business intelligence and compliance reporting

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
A Data Warehouse Framework and Views

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
Architecture of a Three-Tier Data Warehouse

Tier 1: Client workstation

Tier 2: Application server

Tier 3: Database server

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
Architecture of a Two-Tier Data Warehouse

Tier 1: Client workstation

Tier 2: Application & database server

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
Architecture of Web-Based Data Warehousing

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
5 Alternative Data Warehouse Architectures

a. Independent data marts.
b. Data mart bus architecture
c. Hub-and-spoke architecture
d. Centralized data warehouse
e. Federated data warehouse
5 Alternative Data Warehouse Architectures

(a) Independent Data Mart Architectures

Source systems → Staging area → Independent data marts (atomic/summarized data) → End-user access and applications

(b) Data Mart Bus Architecture with Linked Dimensional Data Marts

Source systems → Staging area → Dimensionalized data marts linked by conformed dimensions (atomic/summarized data) → End-user access and applications

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
5 Alternative Data Warehouse Architectures

(c) Hub-and-Spoke Architecture (Corporate Information Factory)

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
5 Alternative Data Warehouse Architectures

(d) Centralized Data Warehouse Architecture

- Source systems
- Staging area
- ETL
- Normalized relational warehouse (atomic/some summarized data)
- End-user access and applications

(e) Federated Architecture

- Existing data warehouses
- Data marts and legacy systems
- Data mapping/metadata
- Logical/physical integration of common data elements
- End-user access and applications

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
# Average Assessment Scores for the Success of the DW Architectures

<table>
<thead>
<tr>
<th></th>
<th>Independent DMs</th>
<th>Bus Architecture</th>
<th>Hub-and-Spoke Architecture</th>
<th>Centralized Architecture (No Dependent DMs)</th>
<th>Federated Architecture</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Information Quality</strong></td>
<td>4.42</td>
<td>5.16</td>
<td>5.35</td>
<td>5.23</td>
<td>4.73</td>
</tr>
<tr>
<td><strong>System Quality</strong></td>
<td>4.59</td>
<td>5.60</td>
<td>5.56</td>
<td>5.41</td>
<td>4.69</td>
</tr>
<tr>
<td><strong>Individual Impacts</strong></td>
<td>5.08</td>
<td>5.80</td>
<td>5.62</td>
<td>5.64</td>
<td>5.15</td>
</tr>
<tr>
<td><strong>Organizational Impacts</strong></td>
<td>4.66</td>
<td>5.34</td>
<td>5.24</td>
<td>5.30</td>
<td>4.77</td>
</tr>
</tbody>
</table>

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
The ETL Process

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
## Sample List of Data Warehousing Vendors

<table>
<thead>
<tr>
<th>Vendor</th>
<th>Product Offerings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Objects (businessobjects.com)</td>
<td>A comprehensive set of BI and data visualization software (now owned by SAP)</td>
</tr>
<tr>
<td>Computer Associates (cai.com)</td>
<td>Comprehensive set of data warehouse (DW) tools and products</td>
</tr>
<tr>
<td>DataMirror (datamirror.com)</td>
<td>DW administration, management, and performance products</td>
</tr>
<tr>
<td>Data Advantage Group (dataadvantagegroup.com)</td>
<td>Metadata software</td>
</tr>
<tr>
<td>Dell (dell.com)</td>
<td>DW servers</td>
</tr>
<tr>
<td>Embarcadero Technologies (embarcadero.com)</td>
<td>DW administration, management, and performance products</td>
</tr>
<tr>
<td>Greenplum (greenplum.com)</td>
<td>Data warehousing and data appliance solution provider (now owned by EMC)</td>
</tr>
<tr>
<td>Harte-Hanks (harte-hanks.com)</td>
<td>Customer relationship management (CRM) products and services</td>
</tr>
<tr>
<td>HP (hp.com)</td>
<td>DW servers</td>
</tr>
<tr>
<td>Hummingbird Ltd. (hummingbird.com)</td>
<td>DW engines and exploration warehouses</td>
</tr>
</tbody>
</table>

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
## Sample List of Data Warehousing Vendors

<table>
<thead>
<tr>
<th>Vendor</th>
<th>Product Offerings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperion Solutions (hyperion.com)</td>
<td>Comprehensive set of DW tools, products, and applications</td>
</tr>
<tr>
<td>IBM InfoSphere (www-01.ibm.com/software/data/infosphere)</td>
<td>Data integration, DW, master data management, Big Data products</td>
</tr>
<tr>
<td>Informatica (informatica.com)</td>
<td>DW administration, management, and performance products</td>
</tr>
<tr>
<td>Microsoft (microsoft.com)</td>
<td>DW tools and products</td>
</tr>
<tr>
<td>Netezza</td>
<td>DW software and hardware (DW appliance) provider (now owned by IBM)</td>
</tr>
<tr>
<td>Oracle (including PeopleSoft and Siebel; oracle.com)</td>
<td>DW, ERP, and CRM tools, products, and applications</td>
</tr>
<tr>
<td>SAS Institute (sas.com)</td>
<td>DW tools, products, and applications</td>
</tr>
<tr>
<td>Siemens (siemens.com)</td>
<td>DW servers</td>
</tr>
<tr>
<td>Sybase (sybase.com)</td>
<td>Comprehensive set of DW tools and applications</td>
</tr>
<tr>
<td>Teradata (teradata.com)</td>
<td>DW tools, DW appliances, DW consultancy, and applications</td>
</tr>
</tbody>
</table>

## Contrasts between the DM and EDW Development Approaches

<table>
<thead>
<tr>
<th>Effort</th>
<th>DM Approach</th>
<th>EDW Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scope</strong></td>
<td>One subject area</td>
<td>Several subject areas</td>
</tr>
<tr>
<td><strong>Development time</strong></td>
<td>Months</td>
<td>Years</td>
</tr>
<tr>
<td><strong>Development cost</strong></td>
<td>$10,000 to $100,000+</td>
<td>$1,000,000+</td>
</tr>
<tr>
<td><strong>Development difficulty</strong></td>
<td>Low to medium</td>
<td>High</td>
</tr>
<tr>
<td><strong>Data prerequisite for sharing</strong></td>
<td>Common (within business area)</td>
<td>Common (across enterprise)</td>
</tr>
<tr>
<td><strong>Sources</strong></td>
<td>Only some operational and external systems</td>
<td>Many operational and external systems</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td>Megabytes to several gigabytes</td>
<td>Gigabytes to petabytes</td>
</tr>
<tr>
<td><strong>Time horizon</strong></td>
<td>Near-current and historical data</td>
<td>Historical data</td>
</tr>
<tr>
<td><strong>Data transformations</strong></td>
<td>Low to medium</td>
<td>High</td>
</tr>
<tr>
<td><strong>Update frequency</strong></td>
<td>Hourly, daily, weekly</td>
<td>Weekly, monthly</td>
</tr>
<tr>
<td><strong>Technology</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hardware</strong></td>
<td>Workstations and departmental servers</td>
<td>Enterprise servers and mainframe computers</td>
</tr>
<tr>
<td><strong>Operating system</strong></td>
<td>Windows and Linux</td>
<td>Unix, Z/OS, OS/390</td>
</tr>
<tr>
<td><strong>Databases</strong></td>
<td>Workgroup or standard database servers</td>
<td>Enterprise database servers</td>
</tr>
<tr>
<td><strong>Usage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number of simultaneous users</strong></td>
<td>10s</td>
<td>100s to 1,000s</td>
</tr>
<tr>
<td><strong>User types</strong></td>
<td>Business area analysts and managers</td>
<td>Enterprise analysts and senior executives</td>
</tr>
<tr>
<td><strong>Business spotlight</strong></td>
<td>Optimizing activities within the business area</td>
<td>Cross-functional optimization and decision making</td>
</tr>
</tbody>
</table>

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*, 4th Edition, Pearson
Essential Differences between Inmon’s and Kimball’s Approaches

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Inmon</th>
<th>Kimball</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Methodology and Architecture</strong></td>
<td>Top-down</td>
<td>Bottom-up</td>
</tr>
<tr>
<td>Overall approach</td>
<td>Enterprise-wide (atomic) data warehouse “feeds” departmental databases</td>
<td>DMs model a single business process, and enterprise consistency is achieved through a data bus and conformed dimensions</td>
</tr>
<tr>
<td>Architecture structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complexity of the method</td>
<td>Quite complex</td>
<td>Fairly simple</td>
</tr>
<tr>
<td>Comparison with established development methodologies</td>
<td>Derived from the spiral methodology</td>
<td>Four-step process; a departure from RDBMS methods</td>
</tr>
<tr>
<td>Discussion of physical design</td>
<td>Fairly thorough</td>
<td>Fairly light</td>
</tr>
<tr>
<td><strong>Data Modeling</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data orientation</td>
<td>Subject or data driven</td>
<td>Process oriented</td>
</tr>
<tr>
<td>Tools</td>
<td>Traditional (entity-relationship diagrams [ERD], data flow diagrams [DFD])</td>
<td>Dimensional modeling; a departure from relational modeling</td>
</tr>
<tr>
<td>End-user accessibility</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td><strong>Philosophy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary audience</td>
<td>IT professionals</td>
<td>End users</td>
</tr>
<tr>
<td>Place in the organization</td>
<td>Integral part of the corporate information factory</td>
<td>Transformer and retainer of operational data</td>
</tr>
<tr>
<td>Objective</td>
<td>Deliver a sound technical solution based on proven database methods and technologies</td>
<td>Deliver a solution that makes it easy for end users to directly query the data and still get reasonable response times</td>
</tr>
</tbody>
</table>

Representation of Data in Data Warehouse

(1) Star Schema  (2) Snowflake Schema

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
# A Comparison between OLTP and OLAP

<table>
<thead>
<tr>
<th>Criteria</th>
<th>OLTP</th>
<th>OLAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>To carry out day-to-day business functions</td>
<td>To support decision making and provide answers to business and management queries</td>
</tr>
<tr>
<td>Data source</td>
<td>Transaction database (a normalized data repository primarily focused on efficiency and consistency)</td>
<td>Data warehouse or DM (a nonnormalized data repository primarily focused on accuracy and completeness)</td>
</tr>
<tr>
<td>Reporting</td>
<td>Routine, periodic, narrowly focused reports</td>
<td>Ad hoc, multidimensional, broadly focused reports and queries</td>
</tr>
<tr>
<td>Resource requirements</td>
<td>Ordinary relational databases</td>
<td>Multiprocessor, large-capacity, specialized databases</td>
</tr>
<tr>
<td>Execution speed</td>
<td>Fast (recording of business transactions and routine reports)</td>
<td>Slow (resource intensive, complex, large-scale queries)</td>
</tr>
</tbody>
</table>

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*, 4th Edition, Pearson
Slicing Operations on a Simple Three-Dimensional Data Cube

A 3-dimensional OLAP cube with slicing/dicing operations

Sales volumes of a specific product on variable time and region

Sales volumes of a specific region on variable time and products

Sales volumes of a specific time on variable region and products

Cells are filled with numbers representing sales volumes
Business Performance Management (BPM) Closed-Loop BPM Cycle

1. STRATEGIZE
   - Mission, values, goals, objectives, incentives, strategy maps

2. PLAN
   - Budgets, plans, forecasts, models, initiatives, targets

INTEGRATED DATA AND METRICS

3. MONITOR/ANALYZE
   - Performance dashboards, reports, analytical tools

4. ACT/ADJUST
   - Interpret, collaborate, assess, decide, act, adjust, track

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
Business Performance Management (BPM) Closed-Loop BPM Cycle

1. **Strategize**
   - *Where do we want to go?*

2. **Plan**
   - *How do we get there?*

3. **Monitor/Analyze**
   - *How are we doing?*

4. **Act and Adjust**
   - *What do we need to do differently?*

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*, 4th Edition, Pearson
Four Perspectives in Balanced Scorecard Methodology

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
Comparison of the Balanced Scorecard and Six Sigma

<table>
<thead>
<tr>
<th>Balanced Scorecard</th>
<th>Six Sigma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategic management system</td>
<td>Performance measurement system</td>
</tr>
<tr>
<td>Relates to the longer-term view of the business</td>
<td>Provides snapshot of business’s performance and identifies measures that</td>
</tr>
<tr>
<td></td>
<td>drive performance toward profitability</td>
</tr>
<tr>
<td>Designed to develop a balanced set of measures</td>
<td>Designed to identify a set of measurements that impact profitability</td>
</tr>
<tr>
<td>Identifies measurements around vision and values</td>
<td>Establishes accountability for leadership for wellness and profitability</td>
</tr>
<tr>
<td>Critical management processes are to clarify vision/</td>
<td>Includes all business processes—management and operational</td>
</tr>
<tr>
<td>strategy, communicate, plan, set targets, align</td>
<td></td>
</tr>
<tr>
<td>strategic initiatives, and enhance feedback</td>
<td></td>
</tr>
<tr>
<td>Balances customer and internal operations without a</td>
<td>Balances management and employees’ roles; balances costs and revenue of</td>
</tr>
<tr>
<td>clearly defined leadership role</td>
<td>heavy processes</td>
</tr>
<tr>
<td>Emphasizes targets for each measurement</td>
<td>Emphasizes aggressive rate of improvement for each measurement,</td>
</tr>
<tr>
<td></td>
<td>irrespective of target</td>
</tr>
<tr>
<td>Emphasizes learning of executives based on feedback</td>
<td>Emphasizes learning and innovation at all levels based on process</td>
</tr>
<tr>
<td></td>
<td>feedback; enlists all employees’ participation</td>
</tr>
<tr>
<td>Focuses on growth</td>
<td>Focuses on maximizing profitability</td>
</tr>
<tr>
<td>Heavy on strategic content</td>
<td>Heavy on execution for profitability</td>
</tr>
<tr>
<td>Management system consisting of measures</td>
<td>Measurement system based on process management</td>
</tr>
</tbody>
</table>

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
Six Sigma

The DMAIC Performance Model

• Define
• Measure
• Analyze
• Improve
• Control

Source: Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson
The Joy of Stats:
200 Countries, 200 Years, 4 Minutes

https://www.youtube.com/watch?v=jbkSRLYSOjo
Python Data Science Handbook in Google Colab

https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/Index.ipynb
Python Data Science Handbook in Google Colab

Table of Contents

Preface

1. IPython: Beyond Normal Python
   • Help and Documentation in IPython
   • Keyboard Shortcuts in the IPython Shell
   • IPython Magic Commands
   • Input and Output History
   • IPython and Shell Commands
   • Errors and Debugging
   • Profiling and Timing Code
   • More IPython Resources

https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/Index.ipynb
2. Introduction to NumPy

- Understanding Data Types in Python
- The Basics of NumPy Arrays
- Computation on NumPy Arrays: Universal Functions
- Aggregations: Min, Max, and Everything In Between
- Computation on Arrays: Broadcasting
- Comparisons, Masks, and Boolean Logic
- Fancy Indexing
- Sorting Arrays
- Structured Data: NumPy's Structured Arrays
3. Data Manipulation with Pandas

• Introducing Pandas Objects
• Data Indexing and Selection
• Operating on Data in Pandas
• Handling Missing Data
• Hierarchical Indexing
• Combining Datasets: Concat and Append
• Combining Datasets: Merge and Join
• Aggregation and Grouping
• Pivot Tables
• Vectorized String Operations
• Working with Time Series
• High-Performance Pandas: eval() and query()
• Further Resources

https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/Index.ipynb
4. Visualization with Matplotlib

- Simple Line Plots
- Simple Scatter Plots
- Visualizing Errors
- Density and Contour Plots
- Histograms, Binnings, and Density
- Customizing Plot Legends
- Customizing Colorbars
- Multiple Subplots
- Text and Annotation
- Customizing Ticks
- Customizing Matplotlib: Configurations and Stylesheets
- Three-Dimensional Plotting in Matplotlib
- Geographic Data with Basemap
- Visualization with Seaborn
- Further Resources

https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/Index.ipynb
5. Machine Learning

• What Is Machine Learning?
• Introducing Scikit-Learn
• Hyperparameters and Model Validation
• Feature Engineering
• In Depth: Naive Bayes Classification
• In Depth: Linear Regression
• In-Depth: Support Vector Machines
• In-Depth: Decision Trees and Random Forests
• In Depth: Principal Component Analysis
• In-Depth: Manifold Learning
• In Depth: k-Means Clustering
• In Depth: Gaussian Mixture Models
• In-Depth: Kernel Density Estimation
• Application: A Face Detection Pipeline
• Further Machine Learning Resources

https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/Index.ipynb
Summary

- Descriptive Analytics II
- Business Intelligence
- Data Warehousing
- Data Integration and the Extraction, Transformation, and Load (ETL) Processes
- Business Performance Management (BPM)
- Performance Measurement
  - Balanced Scorecards
  - Six Sigma
References

• Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson.