

Chapter 1

Introduction to Machine Learning



1.1 Introduction to Machine Learning

The last two decades have seen a quiet but important revolution in computer science. Now more than ever, computers and algorithms are leading to more prosperous and more accurate insights with software that learns from experience and adapts automatically to match the needs of its tasks [1]. Formerly, the programmer decided how the system would work by manually writing the code. Today, we do not write programs but rather collect data consisting of instruction insights, and develop the algorithms changes that manipulate it as necessary to extract patterns and insights. Today, we have programs that can recognize faces and fingerprints, understand speech, translate, navigate, drive a car, recommend movies, and many more [1]. This is possible now because of artificial intelligence (AI) and its fields, mainly machine learning.

Artificial intelligence reflects a computer's ability to recognize patterns to then use and apply those patterns based on available data [2]. Artificial intelligence mimics human cognition by accessing data from a variety of sources and systems to make decisions and learn from their results and patterns [3]. Artificial intelligence was inspired by the human brain, given that computers were once known as "electronic brains" [1]. The human undoubtedly has incredible processing capabilities that humans have long aimed to understand in order to use and create an artificial version known as artificial intelligence. Artificial intelligence has been overgrowing since the Turing Test (originally named the imitation game) was conducted in 1950 by Alan Turing, which suggested that computers do have the ability to think intelligently as artificial entities [3]. Alan Turing's research revolutionized how the world perceived artificial intelligence and its use in daily life [2]. As artificial intelligence has evolved from being a purely academic field, it has become a significant part of many social and economic sectors, including speech recognition, medical diagnosis, vehicles, and voice-activated assistance [3].

Machine learning (ML) is a type of artificial intelligence that branches from computer science [4]. Machine learning is not just a data storing, processing, or training problem; it is instead a means to achieve artificial intelligence by either training it on a dataset or using repeated trials to train a computer program to maximize intelligent performance [1]. Machines are great at making smart decisions because of the enormous datasets. On the other hand, humans are much better at making decisions with limited information. This combination is highly effective in leveraging both human and machine intelligence in creating machine learning models. Combining machine learning and human intelligence provides remarkably high levels of accuracy, leading us to artificial intelligence [3].

1.2 Origin of Machine Learning

Machine learning (ML) is often credited to a psychologist from Cornell University named Frank Rosenblatt, who, based on his theories about the workings of the human nervous system, developed a machine capable of recognizing letters of the alphabet in the 1960s [4]. The machine, called the “perceptron,” converted analog signals into discrete ones, becoming the prototype for modern artificial neural networks (ANNs). Further studies of the structures and learning abilities of neural networks took place in the 1970s and 1980s. Even so, the Novikoff theorem (1962), which states that a perceptron learning algorithm can be converged in a finite number of steps, has become more widely known and credited for machine learning. In 1979, students at Stanford University created a notable invention known as the “Stanford cart,” which could navigate different obstacles in a room [4]. The invention of the Stanford cart is an important part of the history of artificial intelligence and machine learning, as it paved a pathway for robotics research within the area. Today, rover robotic cleaning vacuums use a similar method to avoid obstacles in a room and pick up foreign materials such as dirt.

1.3 Growth of Machine Learning

Machine learning has transformed the twenty-first century through its progress and growth, increasing computer competence in various fields, including the automotive industry, healthcare, commerce, banking, and manufacturing [5]. The first decade of the twenty-first century marked a turning point in machine learning history, which can be attributed to three trends that worked collaboratively [4]. The first trend is **big data**, which refers to a large volume of data that is complicated and requires specialized methods to process [4]. This very large and mostly accessible data includes weather data, business transactions data, medical test results, social media posts, security camera recordings, GPS locations from smartphones, sensor data, and many more. Big data tomorrow will be bigger than today, and with more data,

trained models will get more intelligent [1]. The second trend is the **reduced cost** of parallel computing and memory by distributing the processing of high volumes of data between simple processors [4]. In addition to that, more complex and powerful yet affordable processors like graphics processing units (GPU) are produced. These resources are available for data scientists and organizations today via cloud computing without the need for huge investments in hardware. The reduced costs and increased processing power and storage capability allowed for increased data capturing and storage and more efficient and faster coding and analysis. The third trend is the development of new algorithms of **machine learning** [4], many of which are readily available in open-source communities. The third trend is by far the most important, as it aided in the creation of artificial neural networks that support higher-level functions for data processing [4]. ANNs are crucial algorithms within machine learning (ML), serving the important purpose of solving complex problems.

There is more data than our sensors or brains can handle or process. The information available online today contains massive amounts of digital text and is now so vast that manual processing is impossible. The use of machine learning for this is much more efficient and is known as machine reading. The basic advantage of machine learning is that it can be applied to a wide range of tasks without explicitly being programmed to learn. Using machine learning, we can build systems that are capable of learning and adapting to their environment on their own with minimal supervision and maximal user satisfaction [1].

1.4 How Machine Learning Works

Machine learning works by utilizing many algorithms that make intelligent predictions based on the datasets being used. These datasets can be enormous, consisting of millions of data that cannot be processed by the human mind alone [5]. Machine learning has four variants: supervised, unsupervised, semi-supervised, and reinforcement learning [6]. In order to understand these variants, it is important to understand “labels.” A “label” in machine learning is the dependent variable and is a specified value of the outcome [6]. In supervised learning, labeled datasets are used by machine learning professionals to train algorithms by setting parameters to make accurate predictions about data [3]. Regression is one example of supervised learning [1]. On the other hand, unsupervised learning consists of multiple unlabeled datasets which are used to detect structure and patterns using the algorithm [3]. Data clustering is one example of unsupervised learning, and it is also much faster, as there are fewer labeled data [1]. Semi-supervised learning fits the models to both labeled and unlabeled data [6]. The main goal of semi-supervised learning is to understand how combining labeled and unlabeled data can change learning behavior and the design algorithms that use this combination [7]. Reinforcement learning is a machine learning algorithm that allows machines and software to automatically evaluate optimal behavior in specified contexts for improved efficiency [8]. Reinforced learning is usually used for training complicated artificial

intelligence models to increase automation. All four of these variants are an important component of machine learning outcomes and play a significant role per their learning capabilities [8].

1.5 Machine Learning Building Blocks

Statistics, data mining, analytics, business intelligence, artificial intelligence, and machine learning are concepts, methods, and techniques used to understand data and explore it to find valuable information, relationships, trends, patterns, and anomalies and ultimately to make predictions. In the following section, we will introduce data, its types, and how it is managed and explored. We will also introduce business intelligence and data analytics. Statistics will be covered in detail in Chap. 2, and in Chap. 3, we will introduce data mining and explore machine learning and the different algorithms in more depth.

1.5.1 *Data Management and Exploration*

1.5.1.1 **Data, Information, and Knowledge**

Data and information are comparable but not the same. While many believe that data and information represent the same concept and can be used interchangeably, these two terms are quite distinct. Data are streams of raw, specific, and objective facts or observations generated from events such as business transactions, inventory, or medical examinations. Standing alone, data have no intrinsic meaning. Data is generally broken into two categories, structured and unstructured. Structured data, like patient records, sales transactions, and warehouse inventory, has a predefined format and can be easily processed and analyzed. It is commonly stored and managed in a relational database. Unstructured data, like free text, videos, images, audio files, tweets, and portable medical device outputs, are complex in their form and are more difficult to manage, process, and analyze [9, 10].

Once processed (e.g., filtered, sorted, aggregated, assembled, formatted, or calculated), data becomes endowed with relevance and purpose and is put in context. Data thus turns into information. Information is a cluster of facts that are meaningful and useful to human beings in processes such as making decisions [11, 12]. For instance, patients' IDs, names, dates of birth, home addresses, postal codes, phone numbers, emails, and diagnoses are examples of data that can be collected in a community center, clinic, or hospital, while a bar chart presenting the percentage of patients in different age groups, a pie chart representing the number of patients per type of disease, or a map representing the patients' distribution in a geographic area are examples of information.

Consider a simple example that we can all relate to, how purchases are processed at checkout at a grocery store. Scanning the barcodes of the products at a store generates or accesses data in the form of a product number, a short description of the product, and a price. When these data are processed, an invoice is generated, and the store's inventory is updated. This generated information helps the store determine how much to charge the customer and process the payment. This new information also lets the store manager know how much inventory is left for each product and helps him decide when to order new supplies [13].

In summary, data is the new oil; data is simply a collection of facts. Once data are processed, organized, analyzed, and presented in a way that assists in understanding reality and ultimately making a decision, it is called information. Information is ultimately used to make a decision and take a course of action.

When processed further and internalized by humans, information becomes knowledge. Knowledge can be defined as understanding, awareness, or experience. It can be learned, discovered, perceived, inferred, or understood [10]. In the grocery store example, knowledge would be the awareness of which products sell the most during specific times of the year, which translates into the decision to order additional supplies to avoid out of the stock issue [13].

As we move from data to information and then to knowledge, we see more human contribution and greater value and, traditionally, a decreasing role of technology. Data are easily captured, generated, structured, stored, and transmitted by information and communication technology (ICT). Information, which is data endowed with relevance and purpose [14], requires analysis, a task increasingly being done by technology but also by human mediation and interpretation. Finally, knowledge, valuable information from the human mind, is difficult to capture electronically, structure and transfer and is often tacit and personal to the source [12, 15].

1.5.1.2 Big Data

There is no unique or universal definition of big data, but there is a general agreement that there has been an explosion in data generation, storage, and usage [9]. Big data is a common term used to refer to the massive size structured and unstructured data generated, made available, and being used [16]. These data come from daily business transactions at banks and retailers, for example, from sensors such as security cameras and monitoring systems, the GPS systems on every mobile phone, content posted on social media such as YouTube videos, and from many more ubiquitous sources [13]. Big data in the healthcare field comes from medical devices such as MRI scanners and X-ray machines, sensors such as heart monitors, patient electronic medical and health records, insurance providers' records, doctors' notes, genomic research studies, wearable devices, and many more [17]. An example of big data is what is collected by Fitbit, a manufacturer of wearable activity trackers. In 2018, it was announced that Fitbit had collected 150 billion hours' worth of heart rate data from tens of millions of people from all over the world. These data also

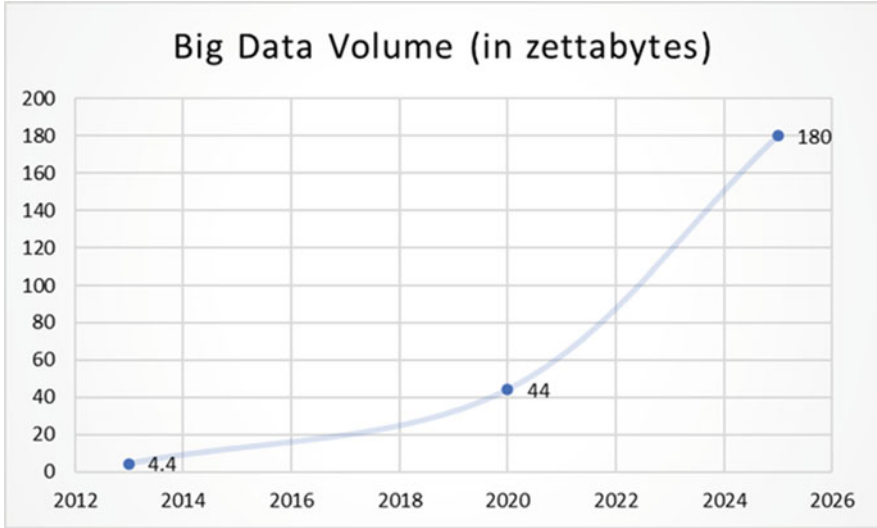


Fig. 1.1 Exponential increase in the volume of big data (actual and projected) based on an IDC study (adapted from [13])

include sex, age, location, height, weight, activity levels, and sleep patterns. Moreover, Fitbit has 6 billion nights' worth of sleep data [18].

There are multiple factors behind the emergence and growth of big data, and they include technological advances in the field of information and communication technology (ICT), where computing power and data storage capacity are continuously increasing while their cost is decreasing. The increased connectivity to the Internet is another major factor. Today, most people have a mobile device, and many modern pieces of equipment are connected to the Internet [13].

Big data is generally characterized by the four Vs: volume, variety, velocity (introduced originally by the Gartner Group in 2001), and veracity (added later by IBM) [19]. Multiple additional Vs were introduced later, including validity, viability, variability, vulnerability, visualization, volatility, and value [9, 19, 20]. Volume is the most defining characteristic of big data. The volume of data generated is increasing exponentially, and new units of measure have been created, such as zettabytes (10^{21}), to accommodate this increasing volume of data. According to IDC, a market-research firm, the data created and copied in 2013 was 4.4 zettabytes, and this number is projected to exponentially increase to 44 zettabytes in 2020 and 180 zettabytes in 2025 (Fig. 1.1) [19, 21]. Examples of large volumes of data are the 20 terabytes (10^{12}) of data produced by Boeing jets every hour and the 1 terabyte of data uploaded on YouTube every 4 minutes [22].

Variety refers to the different forms of big data, such as videos, pictures, social media posts, images from X-ray machines, location data from GPS systems, data from sensors like security devices and wearable wireless health monitors, and many more. **Velocity** refers to the very high speed at which big data are continuously being



Fig. 1.2 Data-to-action value chain

generated, for example, from medical devices and monitors in hospitals' intensive care units or security cameras. Such data must be generated and analyzed in real-time, particularly when the outcome has a direct impact on someone's safety in the case of driverless cars or their financial situation in the case of the stock market. Finally, **veracity** represents the high level of uncertainty and low levels of reliability and truthfulness of big data [9, 10, 13, 19]. Data can be biased, incomplete, or filled with noise, and data scientists and analysts spend more than 60% of their time cleaning data [19]. These characteristics of big data represent challenges for any company or industry. Some of the challenges are technical, such as being able to analyze the large volume of data, which is generated very rapidly and in many different formats. Other challenges may be administrative, such as the reliability of the data [13].

The increasing volume and complexity of data, which is very rapidly generated in different formats, have made it practically impossible for humans to analyze without sophisticated analytics techniques. Therefore, techniques like data analytics, data mining, artificial intelligence, and machine learning are playing an increasing role in transforming data or information into knowledge and helping humans make decisions and take action (Fig. 1.2).

1.5.1.3 OLAP Versus OLTP

A significant amount of data is produced by daily business transactions, be it a purchase of a product, such as an airline ticket or a book; withdrawing money from a bank; admitting a patient to a hospital; generating medical imagery, such as X-rays; updating patient records after a medical examination; and so on [13]. These transactions are managed by transaction processing systems (TPS) or online transaction processing systems (OLTP), which are computerized systems, such as payroll systems, order processing systems, reservations systems, or enterprise resource planning (ERP) systems, that perform and record the transactions that are necessary to conduct a business, such as employee record keeping, payroll, sales order entry, and shipping [11, 23]. At a bank, OLTP can be used to create new accounts, deposit and withdraw funds, process checks, transfer funds to other accounts, withdraw cash, pay bills, calculate and apply fees, and generate a report on all transactions performed during a period of time. OLTP systems function at the operational level of an organization and are mainly responsible for acquiring and storing data related to day-to-day automated business transactions, running everyday real-time analyses, and generating reports [10]. The value of the data generated and maintained by an OLTP system goes beyond supporting an organization's operations and generating reports.

Table 1.1 A comparison between OLTP and OLAP (adapted from Sharda et al. (2015) [10])

Criteria	OLTP	OLAP
Purpose	Support and automate day-to-day business functions and transactions	Support decision-making and provide analytical capabilities and answers to business and management queries
Data sources	Transaction database (a data repository formatted to maximize efficiency and consistency)	Data warehouse or data mart (a data repository formatted to maximize accuracy and completeness)
Reporting	Routine, periodic, narrowly focused reports	Ad-hoc, multidimensional, broadly focused reports and queries
Resource requirement	Ordinary relational databases	Large capacity, specialized databases
Execution speed	Fast (recording business transactions and routine reports)	Relatively slow (resource-intensive, complex, large-scale queries, multiple data sources)

These data, coming from multiple sources or systems, can be further analyzed to support organizational decision-making using online analytical processing (OLAP). OLAP can manipulate and analyze large volumes of data from different perspectives and answer ad-hoc inquiries by executing multidimensional analytical queries [20, 23]. An OLAP system is a computer system with advanced query and analytical functionality, such as ad-hoc and what-if analysis capabilities [20, 24]. At a bank, an OLAP system can be used to predict which customers may quit, an exercise called churn analysis. OLAP can predict which customers are most susceptible to certain new services to develop targeted marketing campaigns instead of blanket or mass marketing, which is more expensive and less efficient and effective. Table 1.1 presents a brief comparison between OLTP and OLAP [10].

1.5.1.4 Databases, Data Warehouses, and Data Marts

Today, most data is stored, organized, manipulated, and managed inside databases. A database is a collection of data formatted and organized into records that facilitates accessing, updating, adding, deleting, and querying those records [25]. A database can be perceived as a collection of files that are viewed as a single storage area of organized data records that are available to a wide range of users [22].

The most common type of database is a relational database, which consists of tables (called relations, hence the name “relational database”) that are connected via relationships (not to be confused with relations or tables). Each table, not different from a spreadsheet, represents an entity of interest for which we collect and manage data, for example, a customer table or a student table. Each table consists of multiple fields related to the entity it represents, such as the customer’s last name, first name, social security number, phone number, and address. An example of an employee relation or table is in Table 1.2.

Table 1.2 An employee table divided into rows (i.e., records) and columns (i.e., fields)

Employee ID	Last name	First name	Department	Home Address	Employment date
1	Gonzales	Valentina	Marketing	55 Main Street	March 17, 2010
2	Singh	Peter	Accounting	12 Lakeshore East	November 5, 2016
3	Watson	Brian	Sales	921 Holmes Avenue	April 7, 2019
4	Jones	Peter	Accounting	6 McGill Crescent	June 24, 2020
5	Smith	Paul	Personnel	1025 Bay Avenue	April 28, 2005

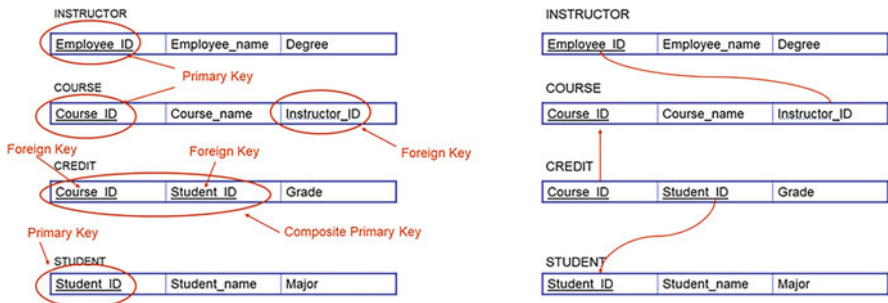


Fig. 1.3 Relationships in a relational database

In a relational database, tables are connected via relationships. Relationships are created by linking primary keys and foreign keys in different tables. In a database, each table has a primary key, which is a field or attribute that is used to uniquely identify each record in the table, such as a student ID, a patient medical health card number, a product code, or a customer phone number. A primary key can consist of multiple fields as long as their combination is unique for every record in the table, such as the combination of a shipment number and product ID. Such a key is called a composite primary key. A table also has foreign keys, which point to primary keys in other tables and confirm the presence of relationships between these tables. Figure 1.3 presents an overview of the relationships in a university database.

Relational databases are designed to quickly access the data for transaction processing (via TPS/OLTP systems), such as admitting a new patient to the hospital or performing a sales transaction. In addition to daily transactions, the database can be queried for occasional reports or information, such as the account balance of a customer; the report can then be used for decision-making, such as providing a line of credit or a loan. However, as the size of a database grows due to day-to-day transactions generating additional new records in the tables, it becomes very time-consuming to generate any analytics using the data. Moreover, analytics or OLAP would slow down the system, making routine transactions handled by the OLTP too slow. Transferring funds between customer bank accounts could take minutes instead of seconds. A solution would be to extract the data from the different

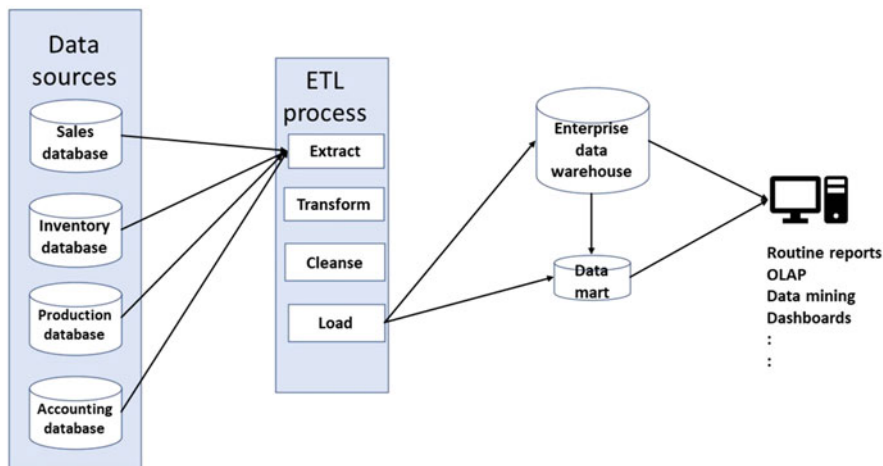


Fig. 1.4 Data warehouse overview (adapted from Sharda et al. (2013) [26])

databases, transform it into an appropriate format, and load it into a special database specifically designed for querying and OLAP. Data that is redundant or has no value would be cleansed. This process is called “extract, transform, and load,” or ETL for short [13]. The databases suitable for querying, OLAP, and decision-making are referred to as “data warehouses” and “data marts.” A data warehouse is a physical repository where current and historical data are specifically organized to provide an enterprise-wide cleansed data in a standardized format [25, 26]. The data in a warehouse is structured to be available in a form ready for OLAP, data mining, querying, reporting, and other decision support applications [26]. A data mart is a data warehouse subset usually focused on a single subject or department [22, 26]. Figure 1.4 presents an overview of a data warehouse.

Data inside a data warehouse or data mart is designed based on the concept of dimensional modeling, where high-volume complex queries are needed. The most common style of dimensional modeling is the star schema. While in an OLTP environment, the database consists of tables representing entities of interest, such as patients and their attributes (name, phone, address, etc.), the star schema has a fact table with a large number of attributes (mainly numbers) that are most needed for analysis and queries, while the rest of the valuable data are stored in attached dimension tables [24, 26, 27]. Figure 1.5 provides a visual representation of an OLTP and an OLAP-based database structure.

While an OLTP system is designed for operational purposes and thus it is detailed, continuously and easily updated, has very current data, has to be always available, and is designed for transactional speed, an OLAP-based system is more informational and has summary data, is not continuously updated, has mainly historical and integrated data, and is designed for complex queries and analytics [24, 26, 27].

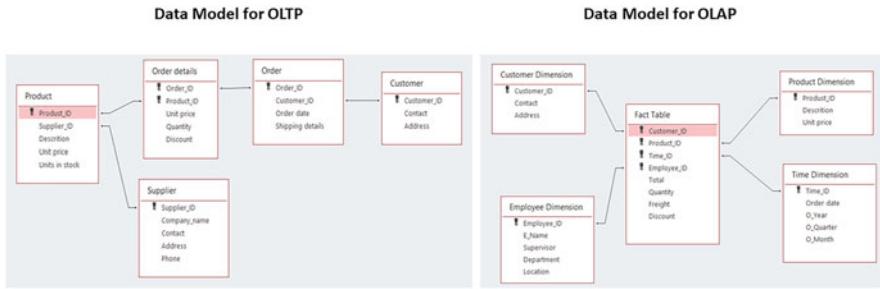


Fig. 1.5 Data models for OLTP and OLAP (adapted from Mailvaganam (2007) [27])

To perform the analytical processing, a data cube is created, which is a multidimensional data structure that is generated out of the star schema and allows for fast analysis of data. The cube is multidimensional but is commonly represented as three-dimensional for ease of viewing and understanding. Each side in a cube represents a dimension, such as a patient, procedure, or time, and the cells are populated with data from the fact table. The cube is optimally designed for common OLAP operations, such as filtering, slicing, dicing, drilling up and down, rolling up, and pivoting [24, 26, 27], which will be explored next.

1.5.1.5 Multidimensional Analysis Techniques

Data to be reported can be manipulated in many cases with simple arithmetic and statistical operations, such as summing up (e.g., total sales in a year), counting (e.g., the number of sales transactions), calculating the mean (e.g., the average profit of sales), filtering (e.g., extracting names of customers in a certain region who made the highest purchases), sorting, ranking, and so on. To extract the data from multiple tables in a relational database of a TPS system, one can issue an SQL query command that pulls out the data from multiple tables by performing a “join” operation (i.e., joining related data from different tables). To perform OLAP on a multidimensional data structure, similar to a cube in a data warehouse, several operations or techniques may be needed, such as slicing, dicing, and pivoting [9, 24, 28].

For simplicity, assume that we have a three-dimensional dataset of sales, where the dimensions are product, region, and year, which can be represented as a cube, where each axis is a dimension, and the cells contain sales data in thousands of dollars (Fig. 1.6). In this figure, we find that the sale of chairs in QC in 2021 was worth \$110,000.

Fig. 1.6 Example of an OLAP cube containing sales figures of three products in three regions over a period of 3 years

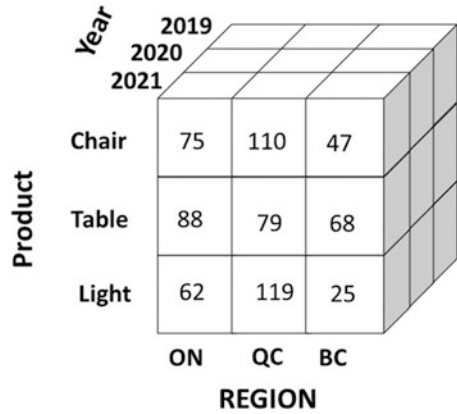
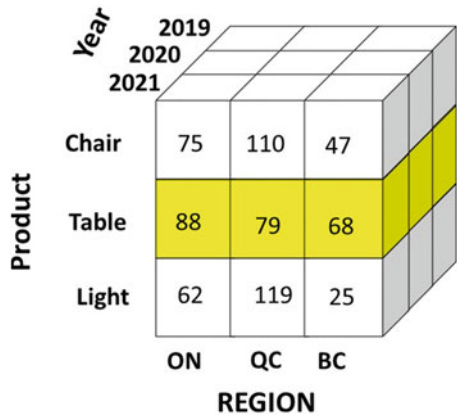


Fig. 1.7 Representation of slicing data to extract the sales of tables in all regions and years



1.5.1.5.1 Slicing and Dicing

Slicing and dicing operations are used to make large amounts of data easier to understand and work with. Slicing is a method to filter a large dataset into smaller datasets of interest while dicing these datasets creates even more granularly defined datasets [9]. Slicing is taking a single slice out of the cube, representing one dimension, showing, for example, the sales of tables for each region and year (Fig. 1.7).

Another example of slicing is in Table 1.3, where the sales of each product are summed up for all regions and years.

Dice is a slice on more than two dimensions of the cube [28]. Dicing is putting multiple side-by-side members from a dimension on an axis with multiple related members from a different dimension on another axis, allowing the viewing and analysis of the interrelationship among different dimensions [24]. Two examples of

Table 1.3 Example of slicing

Sales for all regions and years	
Product	Sales
Chair	\$694,000
Table	\$710,000
Light	\$575,000
Total sales	\$1,979,000

dicing are depicted in Table 1.4, showing the sales of all products per region per year and sales per region per product for all years.

1.5.1.5.2 Pivoting

A pivot table is a cross-tabulated structure (crosstab) that displays aggregated and summarized data based on the ways the columns and rows are sorted. Pivoting means swapping the axes or exchanging rows with columns and vice versa or changing the dimensional orientation of a report [9, 24, 28] (Table 1.5).

1.5.1.5.3 Drill-Down, Roll-Up, and Drill-Across

Drilling down or rolling up is where the user navigates among levels of the data ranging from the most summarized (roll-up) to the most detailed (drill-down) [28] and happens when there is a multilevel hierarchy in the data (e.g., country, province, city, neighborhood) and the users can move from one level to another [24]. Figure 1.8 shows an example of drilling down on the product dimension. When you roll up, the key data, such as sales, are automatically aggregated, and when you drill down, the data are automatically disaggregated [9].

Drilling across is a method where you drill from one dimension to another, but where the drill-across path must be defined [24]. Figure 1.9 shows an example of a drill-across from the store CA to the product dimension.

1.5.2 The Analytics Landscape

Analytics is the science of analysis—using data for decision-making [26]. Analytics involves the use of data, analysis, and modeling to arrive at a solution to a problem or to identify new opportunities. Data analytics can answer questions such as (1) what has happened in the past and why, (2) what could happen in the future and with what certainty, and (3) what actions can be taken now to control events in the future [9, 10].

Data analytics have traditionally fallen under the umbrella of a larger concept called “business intelligence,” or BI. BI has been defined as “the integration of data

Table 1.4 Dicing for region/year (left) and region/product (right)

	Year			Total	Region	Product			Total
	2019	2020	2021			Chair	Table	Light	
Region									
ON	\$220,000	\$229,000	\$225,000	\$674,000	ON	\$240,000	\$239,000	\$190,000	\$669,000
QC	\$290,000	\$303,000	\$299,000	\$892,000	QC	\$249,000	\$246,000	\$201,000	\$696,000
BC	\$130,000	\$143,000	\$140,000	\$413,000	BC	\$205,000	\$225,000	\$184,000	\$614,000
Total	\$640,000	\$675,000	\$664,000	\$1,979,000	Total	\$694,000	\$710,000	\$575,000	\$1,979,000

Table 1.5 Pivoting region and year

Region	Year			Total	Year	Region			Total
	2019	2020	2021			ON	QC	BC	
ON	\$220,000	\$229,000	\$225,000	\$674,000	2019	\$220,000	\$290,000	\$130,000	\$640,000
QC	\$290,000	\$303,000	\$299,000	\$892,000	↔	\$229,000	\$303,000	\$143,000	\$675,000
BC	\$130,000	\$143,000	\$140,000	\$413,000	2021	\$225,000	\$299,000	\$140,000	\$664,000
Total	\$640,000	\$675,000	\$664,000	\$1,979,000	Total	\$674,000	\$892,000	\$413,000	\$1,979,000

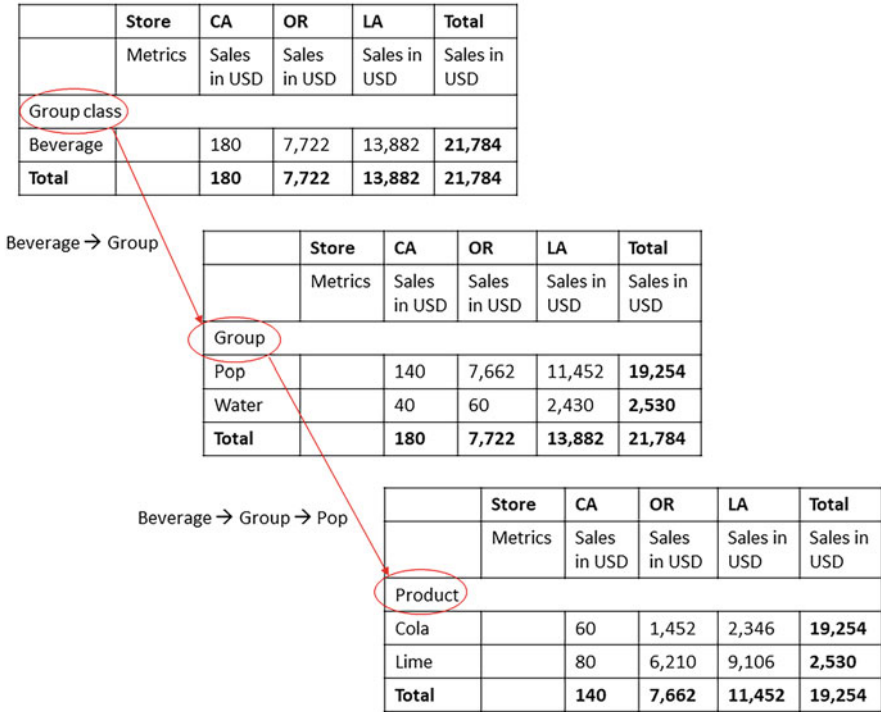


Fig. 1.8 Drilling down (adapted from Ballard et al. (2012) [24])

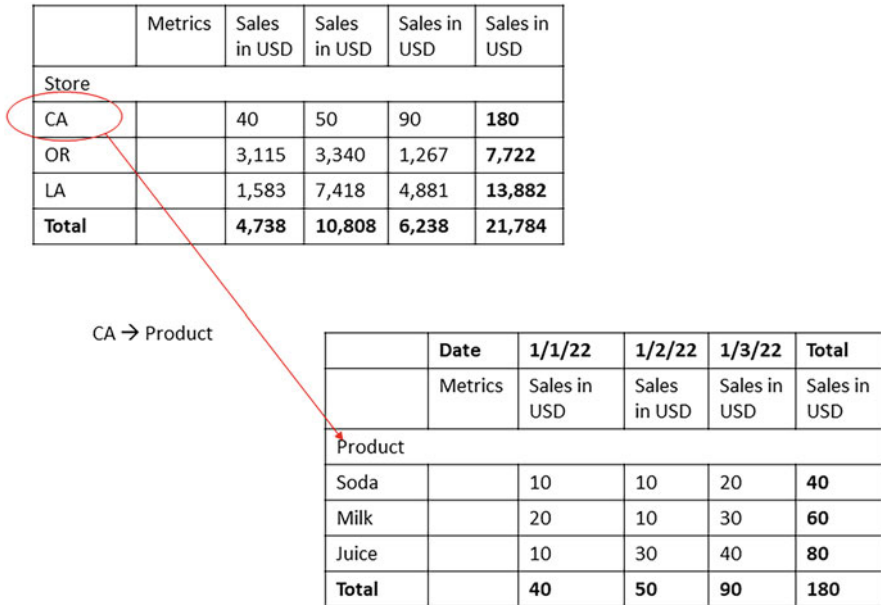


Fig. 1.9 Drilling across (adapted from Ballard et al. (2012) [24])

from disparate source systems to optimize business usage and understanding through a user-friendly interface” [29] and as “the concepts and methods to improve business decision-making by using fact-based support systems” [30]. BI is a conceptual framework for decision support that combines a system architecture, databases and data warehouses, analytical tools, and applications [22]. BI is a mature concept that applies to many fields, despite the presence of the word “business.” While remaining a quite common term, BI is slowly being replaced by the term “analytics,” sometimes referring to the same thing. The major objective of BI is to enable interactive access to data and data manipulation, and to provide end users (e.g., managers, professionals) with the capacity to perform analysis for decision-making. BI analyzes historical and current data and transforms it into information and valuable insights (and knowledge), which lead to more informed and evidence-based decision-making [10]. BI has been very valuable in applications such as customer segmentation in marketing, fraud detection in finance, demand forecasting in manufacturing, and risk factor identification and disease prevention and control in healthcare. BI uses a set of metrics to measure past performance and report a set of indicators that can guide decision-making; it involves a set of methods such as querying structured datasets and reporting the findings, using dashboards, automated monitoring of critical situations, online analytical processing (OLAP) using cubes, slice and dice, and drilling. BI is essentially reactive and performed with much human involvement [13].

Analytics, alternately, are more proactive and can be performed automatically by a set of algorithms (e.g., data mining and machine learning algorithms). Analytics access structured data (e.g., product code, quantity sold, and current inventory level) and unstructured data (e.g., free text describing the product or pictures of the product); they describe what happened in the past, such as how many units of a certain product were sold last year (descriptive analytics); predict what will (most likely) happen in the future, such as how many units we expect to sell next year (predictive analytics); or even prescribe what actions we should take to have certain outcomes in the future (prescriptive analytics), such as what quantity of the product we should order and when. Analytics analyze trends, recognize patterns, and possibly prescribe actions for better outcomes, and they use a multitude of methods, such as predictive modeling, data mining, text mining, statistical analysis, simulation, and optimization [13].

Some sources offer a distinction between BI and analytics using a spectrum of analytics capabilities. BI is traditional and mature and looks at the present and historical data to describe the current state of a business. It uses basic calculations to provide answers. This functionality is compatible with what is referred to as “descriptive analytics” and is at the lower end of the spectrum. Analytics, on the other hand, mines data to predict where the business is heading and prescribes actions to maximize beneficial outcomes. It uses mathematical models to determine attributes and offer predictions. These functionalities are referred to as “predictive” and “prescriptive analytics” and fall on the higher end of the analytics spectrum [13, 31]. Having clarified to a certain extent the difference between BI and analytics, we will refrain from using the term BI and rely instead on the analytics taxonomy:

descriptive, diagnostic, predictive, and prescriptive analytics, which will be described in detail below.

1.5.2.1 Types of Analytics (Descriptive, Diagnostic, Predictive, Prescriptive)

Analytics are of four types: descriptive, diagnostic, predictive, and prescriptive. These types have increasing difficulty and complexity levels and provide increasing value to the users (Fig. 1.10).

1.5.2.1.1 Descriptive Analytics

Descriptive analytics query past or current data and report on what happened (or is happening). Descriptive analytics display indicators of past performance to assist in understanding successes and failures and provide evidence for decision-making; for instance, decisions related to the delivery of quality care and optimization of performance need to be based on evidence [13].

Using descriptive analytics, such as reports and data visualization tools (e.g., dashboards), end users can look retrospectively into past events; draw insight across different units, departments, and, ultimately, the entire organization; and collect evidence that is useful for an informed decision-making process and evidence-based actions. At the initial stages of analysis, descriptive analytics provide an understanding of patterns in data to find answers to the “What happened?” questions,

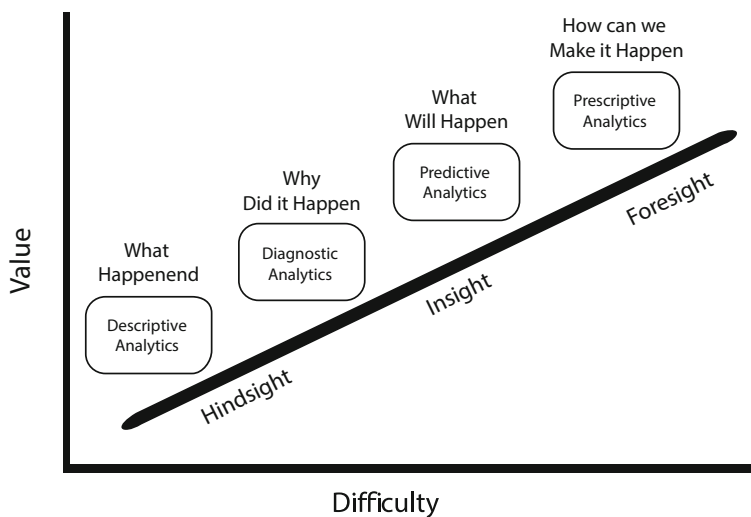


Fig. 1.10 Types of analytics, the value they provide, and their level of difficulty (adapted from Rose Business Technologies [32])

for example, “Who are our best customers in terms of sales volume?” and “What are our least selling products?” Descriptive statistics, such as measures of central tendency (mean, median, and mode) and measures of dispersion (minimum, maximum, range, quartiles, and standard deviations), as well as distribution of variables (e.g., histograms), are used in descriptive analytics [13].

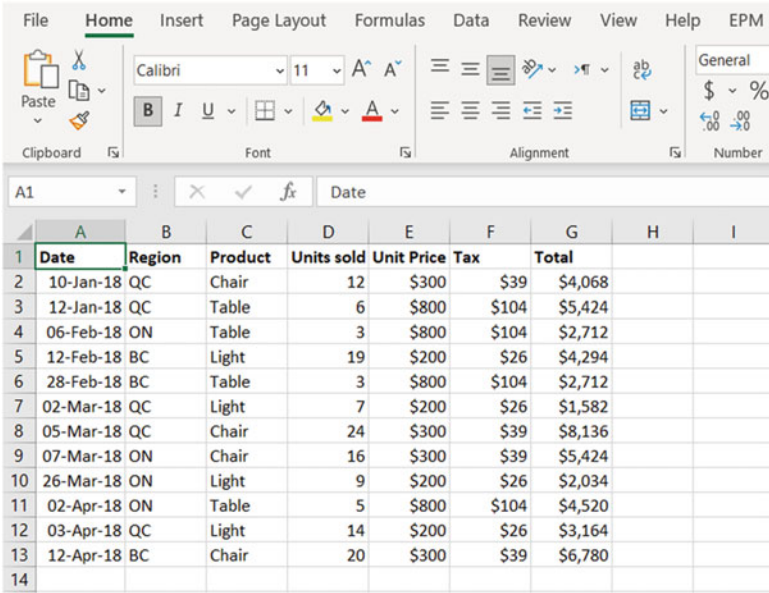
Descriptive analytics can quantify events and report on them and are a first step in turning data into actionable insights. Descriptive analytics, for example, can help with population health management tasks, such as identifying how many patients are living with diabetes, benchmarking outcomes against government expectations, or identifying areas for improvement in clinical quality measures or other aspects of care [33]. Descriptive analytics considers past data analysis to make decisions that help us achieve current and future goals. Statistical analysis is the main “tool” used to perform descriptive analytics; it includes descriptive statistics that provide simple summaries, including graphics analysis, measures of central tendencies (e.g., frequency graphs, average/mean, median, mode), or measures of data variation or dispersion (e.g., standard deviation) [13].

Surveys, interviews, focus groups, web metrics data (e.g., number of hits on a webpage, number of visitors to a page), app metrics data (e.g., number of minutes spent using a feature), and health data stored in electronic records can be the source for all analytics, including descriptive analytics. Media companies and social media platforms (e.g., Facebook) use descriptive analytics to measure customer engagement; managers in hospitals can use descriptive analytics to understand the average wait times in the emergency room (ER) or the number of available beds. Descriptive analytics allow us to access information needed to make *actionable decisions* in the workplace. They allow decision-makers to *explore* trends in data (why do we have long lines in the ER?), to *understand* the “business” environment (who are the patients coming to the ER?), and to possibly *infer an association* (i.e., a correlation) between an outcome and some other variables (patients with the chronic obstructive pulmonary disease tend to have more visits to the ER) [13].

Reports are the main output in descriptive analytics, where findings are presented in charts (e.g., a bar graph or pie chart), summary tables, and most interestingly, pivot tables. A pivot table is a table that summarizes data originating from another table and provides users with the functionality to sort, average, sum, and group data in a meaningful way [13] (Fig. 1.11, 1.12 and 1.13).

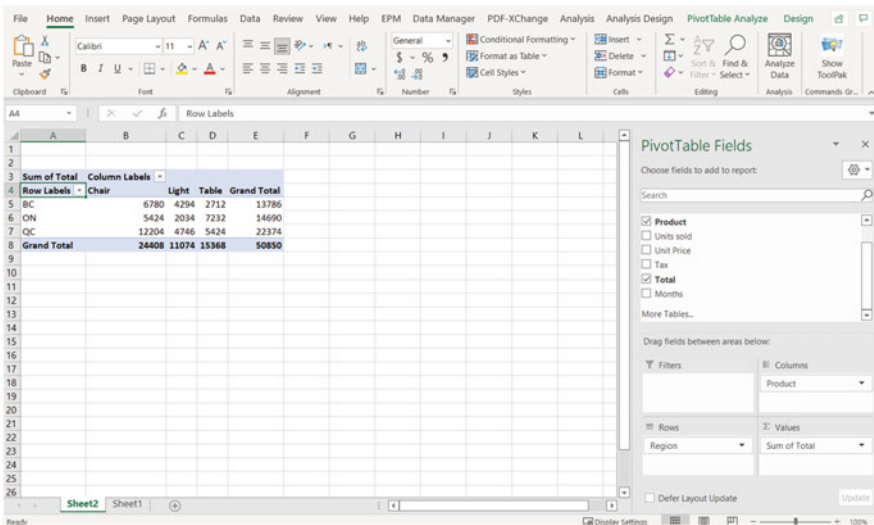
1.5.2.1.2 Diagnostic Analytics

Descriptive analytics give us insight into the past but do not answer the question, “Why did it happen?” Diagnostic analytics aims to answer that type of question. They focus on enhancing processes by identifying why something happened and what the relationships are between the event and other variables that could constitute its causes [34]. They involve trend analysis, root cause analysis [35], cause and effect analysis [36, 37], and cluster analysis [38]. They are exploratory and provide users with interactive data visualization tools [39]. An organization can monitor its performance indicators through diagnostic analysis.



	A	B	C	D	E	F	G	H	I
1	Date	Region	Product	Units sold	Unit Price	Tax	Total		
2	10-Jan-18	QC	Chair	12	\$300	\$39	\$4,068		
3	12-Jan-18	QC	Table	6	\$800	\$104	\$5,424		
4	06-Feb-18	ON	Table	3	\$800	\$104	\$2,712		
5	12-Feb-18	BC	Light	19	\$200	\$26	\$4,294		
6	28-Feb-18	BC	Table	3	\$800	\$104	\$2,712		
7	02-Mar-18	QC	Light	7	\$200	\$26	\$1,582		
8	05-Mar-18	QC	Chair	24	\$300	\$39	\$8,136		
9	07-Mar-18	ON	Chair	16	\$300	\$39	\$5,424		
10	26-Mar-18	ON	Light	9	\$200	\$26	\$2,034		
11	02-Apr-18	ON	Table	5	\$800	\$104	\$4,520		
12	03-Apr-18	QC	Light	14	\$200	\$26	\$3,164		
13	12-Apr-18	BC	Chair	20	\$300	\$39	\$6,780		
14									

Fig. 1.11 Example of a data sheet in Microsoft Excel



Sum of Total	Column Labels	Light	Table	Grand Total	
BC	Chair	6780	4294	2712	13786
ON	Chair	5424	2034	7232	14690
QC	Chair	12204	4746	5424	22374
Grand Total		24408	11074	15368	50850

Fig. 1.12 Example of a pivot table in Microsoft Excel that summarizes data from Fig. 1.11. On the right, we can notice the pivot table fields that allow users to control which summaries are being computed and displayed

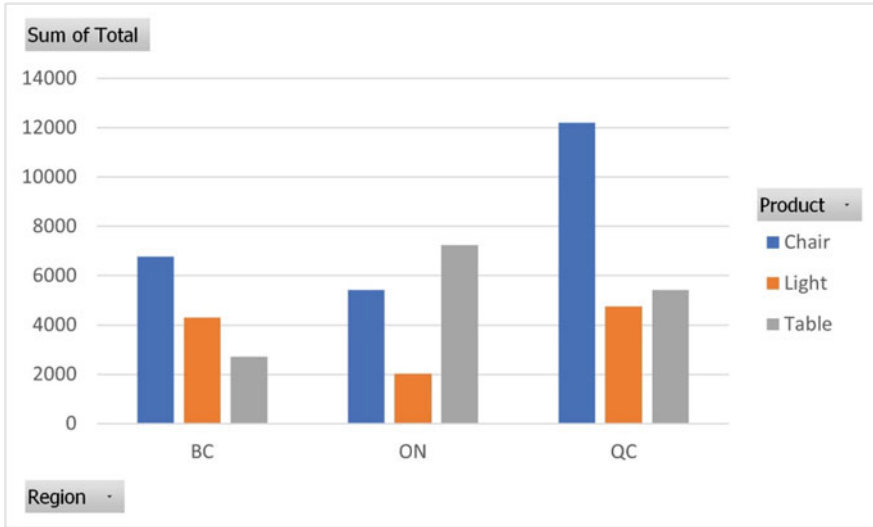


Fig. 1.13 Example of a column chart in Microsoft Excel that visualizes data from Fig. 1.12

1.5.2.1.3 Predictive Analytics

Predictive analysis uses past data to create a model that answers the question, “What will happen?”; it analyzes trends in historical data and identifies what is likely to happen in the future. Using predictive analytics, users can prepare plans and proactively implement corrective actions in advance of the occurrence of an event [39]. Some of the techniques used are what-if analysis, predictive modeling [40–42], machine learning algorithms [43–45], and neural network algorithms [46, 47]. Predictive analytics can be used for forecasting and resource planning. Predictive analytics share many basic concepts and techniques, like algorithms, with machine learning, which is covered in detail later in this textbook.

1.5.2.1.4 Prescriptive Analytics

While predictive analytics estimate what may happen in the future, prescriptive analytics goes a step further by prescribing a certain action plan to address the problems revealed by diagnostic analytics and increase the likelihood of the occurrence of the desired outcome (which may not have been forecasted by predictive analytics) [39, 48–50]. Prescriptive analytics encompasses *simulating*, evaluating several *what-if scenarios*, and advising how to maximize the likelihood of the occurrence of desired outcomes. Some of the techniques used in prescriptive analytics are graph analysis, simulation [51–53], stochastic optimization [54–56], and nonlinear programming [57–59]. Prescriptive analytics is beneficial for advising a course of action to reach a desirable goal.

Evolution of Analytics since the 1980s				
	DESCRIPTIVE ANALYTICS (FOUNDATIONAL)	DIAGNOSTIC ANALYTICS (OPERATIONAL)	PREDICTIVE ANALYTICS (INSIGHTFUL)	PRESCRIPTIVE ANALYTICS (STRATEGIC)
QUESTION	1- What happened in the past	2- what is happening now?	3- Why did it happen and what are the relationships?	4- What will happen in the future?
PROCESS FOCUS	-Reporting	-Measuring KPIs	Trend Analysis Root Cause Analysis Cluster Analysis	Forecasting, Risk Management, Prediction Probability assessment
TOOLS AND TECHNIQUES	-Static and Interactive Reports	-Dashboard -Performance scoreboards	Data Mining Modeling OLAP, Decision Trees	Data Mining Modeling OLAP, Decision Trees
	Retrospective		Prospective	

Fig. 1.14 Analytics: questions, focus, and tools (adapted from Podolak [60])

Prescriptive analytics go beyond prediction to prescribe an optimal course of action to reach a certain goal based on predictions of future events. A simple example would be an app that predicts the duration of a journey from a current location to certain destinations; if the app is equipped with prescriptive analytics, then it can prescribe the shortest path to reach the destination after comparing several alternative routes [13] (Fig. 1.14).

1.6 Conclusion

Machine learning has proven itself to be a sustainable and useful technology in today’s world, and its use is increasing every single day. Everything from smart devices to sophisticated automated systems such as self-driving cars uses machine learning in order to operate. Our progressively complex world is better understood with machine learning because we are currently exposed to more information than ever before and it will only continue growing [1]. In this chapter, we introduced the concept of machine learning and its origins, applications, and building blocks. In the following chapters, we elaborate more on the concept and explore in depth its different algorithms.

1.7 Key Terms

1. Machine learning
2. Artificial intelligence
3. Parallel computing
4. Distributed computing

5. Graphics processing units (GPU)
6. Big data
7. Transaction processing systems (TPS)
8. Online transaction processing systems (OLTP)
9. Online analytical processing (OLAP)
10. Data variety
11. Data velocity
12. Data veracity
13. Databases
14. Data warehouses
15. Data marts
16. Data slicing
17. Data dicing
18. Analytics
19. Descriptive analytics
20. Diagnostic analytics
21. Predictive analytics
22. Prescriptive analytics

1.8 Test Your Understanding

1. Write a definition of analytics.
2. How are descriptive analytics different than diagnostic analytics?
3. How are diagnostic analytics different than predictive analytics?
4. What is data slicing?
5. When do we use data dicing?
6. Which system is focused on daily business processes: OLTP or OLAP?
7. Enumerate five advantages of big data in healthcare.
8. Choose a sector of society and specify five advantages of the use of AI and machine learning in that sector.
9. Choose a sector of society and specify five disadvantages of the use of AI and machine learning in that sector.
10. Can AI be a source of bias? How? Search for examples in the literature.

1.9 Read More

1. Biswas, R. (2021). Outlining Big Data Analytics in Health Sector with Special Reference to Covid-19. *Wirel Pers Commun*, 1–12. <https://doi.org/10.1007/s11277-021-09446-4>
2. Clark, C. R., Wilkins, C. H., Rodriguez, J. A., Preininger, A. M., Harris, J., DesAutels, S., Karunakaram, H., Rhee, K., Bates, D. W., & Dankwa-Mullan, I. (2021). Health Care Equity in the Use of Advanced Analytics and Artificial

- Intelligence Technologies in Primary Care. *J Gen Intern Med*, 36(10), 3188–3193. <https://doi.org/10.1007/s11606-021-06846-x>
3. El Morr, C., & Ali-Hassan, H. (2019). *Analytics in Healthcare: A Practical Introduction*. Springer.
 4. IBM. (2022). What are healthcare analytics? IBM. Retrieved May, 10, 2022 from <https://www.ibm.com/topics/healthcare-analytics>
 5. Khalid, S., Yang, C., Blacketer, C., Duarte-Salles, T., Fernández-Bertolín, S., Kim, C., Park, R. W., Park, J., Schuemie, M. J., Sena, A. G., Suchard, M. A., You, S. C., Rijnbeek, P. R., & Reps, J. M. (2021). A standardized analytics pipeline for reliable and rapid development and validation of prediction models using observational health data. *Comput Methods Programs Biomed*, 211, 106,394. <https://doi.org/10.1016/j.cmpb.2021.106394>
 6. Lopez, L., Chen, K., Hart, L., & Johnson, A. K. (2021). Access and Analytics: What the Military Can Teach Us About Health Equity. *Am J Public Health*, 111(12), 2089–2090. <https://doi.org/10.2105/ajph.2021.306535>
 7. Moreno-Fergusson, M. E., Guerrero Rueda, W. J., Ortiz Basto, G. A., Arevalo Sandoval, I. A. L., & Sanchez-Herrera, B. (2021). Analytics and Lean Health Care to Address Nurse Care Management Challenges for Inpatients in Emerging Economies. *J Nurs Scholarsh*, 53(6), 803–814. <https://doi.org/10.1111/jnu.12711>
 8. Mukherjee, S., Frimpong Boamah, E., Ganguly, P., & Botchwey, N. (2021). A multilevel scenario based predictive analytics framework to model the community mental health and built environment nexus. *Sci Rep*, 11(1), 17,548. <https://doi.org/10.1038/s41598-021-96801-x>
 9. Qiao, S., Li, X., Olatosi, B., & Young, S. D. (2021). Utilizing Big Data analytics and electronic health record data in HIV prevention, treatment, and care research: a literature review. *AIDS Care*, 1–21. <https://doi.org/10.1080/09540121.2021.1948499>

1.10 Lab

All instructions will be for Windows users; Mac users can follow overall the same instructions.

1.10.1 Introduction to R

R is an open-source integrated development environment (IDE) used for statistical analysis. This section describes step-by-step instructions to download and install R v4.1.0 and RStudio IDE v1.4.1717.

R can be downloaded and installed from the following location: <https://www.r-project.org/>. Below are instructions for R's download and installation.

1. Go to the following mirror location and download R (The Comprehensive R Archive Network (sfu.ca)). For Windows users, click on “Download R for Windows”; for other operating systems, click on the corresponding link (Fig. 1.15).
2. While we will demonstrate the installation for Windows, the installation for macOS is similar. For Windows, click on the “install R for the first time” link as shown in Fig. 1.16:
3. Click on the “Download R 4.1.0 Windows” link (Fig. 1.17):
4. R-4.1.0-win.exe will be installed into the Downloads folder. Click on the R-4.1.0-win.exe file to start the installation and continue by clicking the “Next” button to complete the installation (Fig. 1.18).
5. If a shortcut for R was not automatically created on your desktop, you can always create one. On Windows, go to the following location C:\Program Files\R-4.1.0 (on a Mac, go to the Applications folder) (Fig. 1.19), right click on R.exe, and choose “Create Shortcut” (or “Make Alias” for Mac users). A new shortcut will be created; move it to your desktop. This will enable you to launch R easily from the desktop.
6. Double-click on the “R” icon and launch the R software; a new command prompt appears (Fig. 1.20):

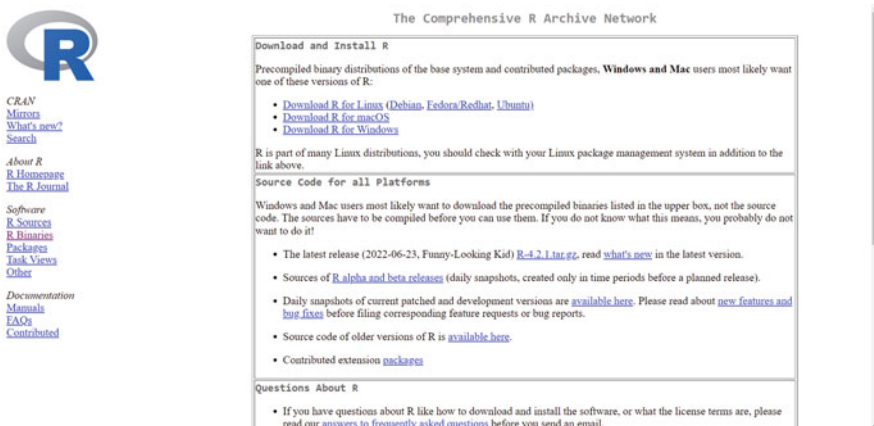


Fig. 1.15 R Installation for Windows users and Mac users

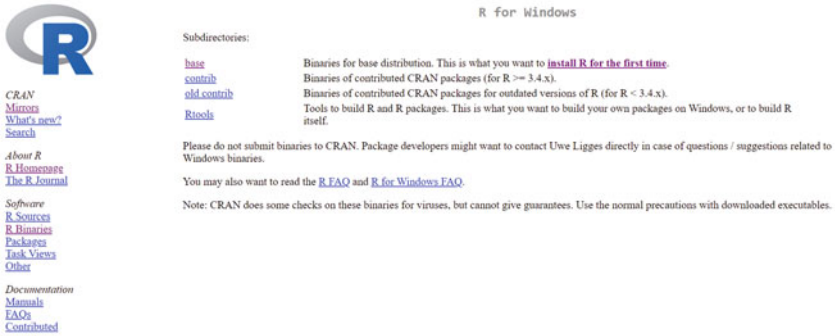


Fig. 1.16 First-time R installation



Fig. 1.17 R v4.1.0 for Windows installation

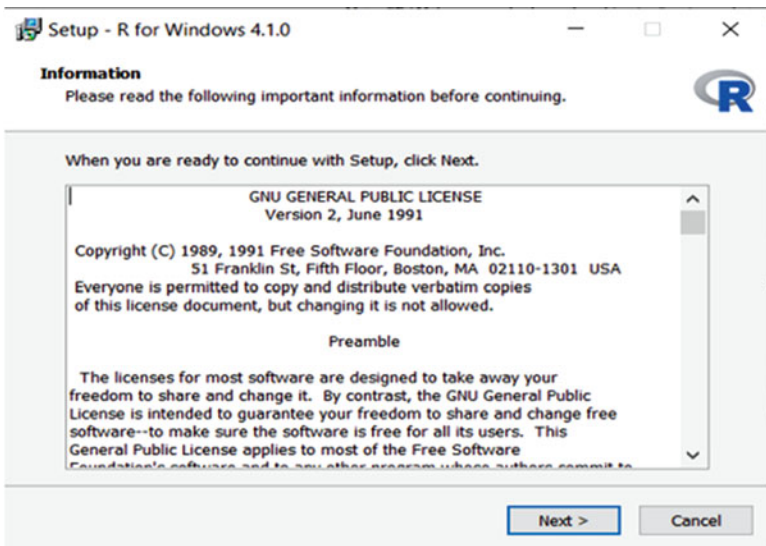


Fig. 1.18 R setup for Windows user

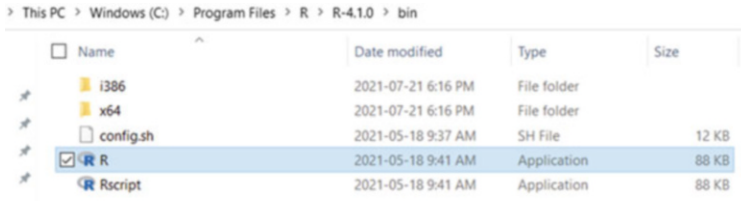


Fig. 1.19 R installation local directory



Fig. 1.20 Launching R application

1.10.2 Introduction to RStudio

RStudio v1.4.1717 is the IDE for the R language. It includes a workspace for coding, debugging, plotting, etc. It can be installed from the following location: Download the RStudio IDE—RStudio. Below are instructions for RStudio’s download and installation.

1.10.2.1 RStudio Download and Installation

1. Download RStudio: Click on “Download RStudio for Windows.” RStudio is available for Mac users as well on the same webpage (Fig. 1.21).
2. Double-click on “RStudio-1.4.1717.exe” and click on “Setup RStudio v1.4.1717”; the installation will start (Fig. 1.22).

RStudio Desktop 1.4.1717 - Release Notes

1. Install R. RStudio requires R 3.0.1+.
2. Download RStudio Desktop. Recommended for your system:



Requires Windows 10 (64-bit)

Fig. 1.21 Installing RStudio Desktop 1.4.1717

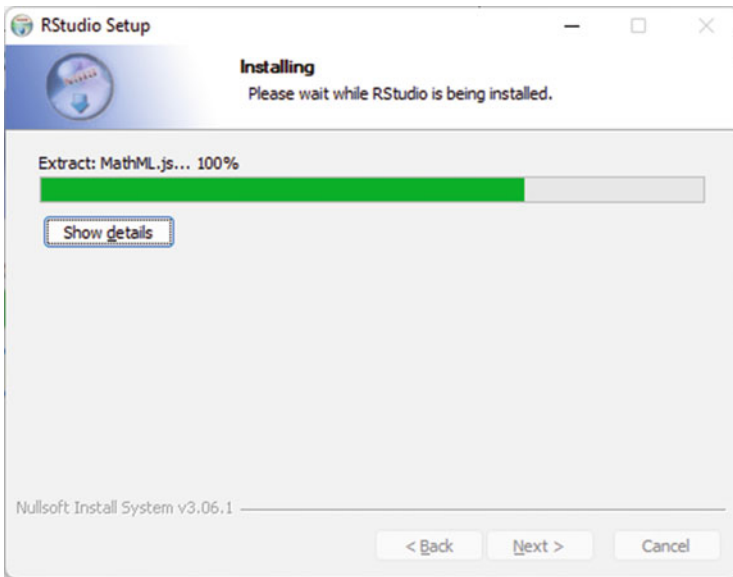


Fig. 1.22 RStudio setup

3. If a shortcut for RStudio was not automatically created on your desktop, you can always create one. On Windows, go to the following location `c:\Program files\RStudio\Bin` (go to the Applications folder if you are using macOS) (Fig. 1.23).
4. After launching the RStudio application, its IDE will appear as shown in Fig. 1.24:

> Windows (C:) > Program Files > RStudio > bin

Name	Date modified	Type	Size
libEGL.dll	2020-03-31 12:38 AM	Application extens...	24 KB
libGLESV2.dll	2020-03-31 12:38 AM	Application extens...	3,496 KB
msvcp140.dll	2021-05-24 6:45 PM	Application extens...	614 KB
msvcp140_1.dll	2021-05-24 6:45 PM	Application extens...	31 KB
msvcp140_2.dll	2021-05-24 6:45 PM	Application extens...	202 KB
opengl32sw.dll	2016-06-14 8:00 AM	Application extens...	20,433 KB
Qt5Core.dll	2021-05-24 8:20 PM	Application extens...	6,021 KB
Qt5Gui.dll	2020-03-31 12:38 AM	Application extens...	6,364 KB
Qt5Network.dll	2020-03-31 12:38 AM	Application extens...	1,297 KB
Qt5Positioning.dll	2020-03-31 3:03 PM	Application extens...	316 KB
Qt5PrintSupport.dll	2020-03-31 12:38 AM	Application extens...	317 KB
Qt5Qml.dll	2020-03-31 12:11 PM	Application extens...	3,933 KB
Qt5Quick.dll	2020-03-31 12:11 PM	Application extens...	3,709 KB
Qt5QuickWidgets.dll	2020-03-31 12:11 PM	Application extens...	78 KB
Qt5SerialPort.dll	2020-03-31 12:43 PM	Application extens...	85 KB
Qt5Svg.dll	2020-03-31 11:27 AM	Application extens...	331 KB
Qt5WebChannel.dll	2020-03-31 1:18 PM	Application extens...	121 KB
Qt5WebEngineCore.dll	2020-04-02 2:37 PM	Application extens...	78,152 KB
Qt5WebEngineWidgets.dll	2020-04-02 2:37 PM	Application extens...	227 KB
Qt5Widgets.dll	2020-03-31 12:38 AM	Application extens...	5,463 KB
QtWebEngineProcess	2020-04-02 2:37 PM	Application	25 KB
rpostback	2021-05-24 7:38 PM	Application	5,285 KB
rsession	2021-05-24 7:58 PM	Application	24,484 KB
rsinverse	2021-05-24 7:46 PM	Application	6,266 KB
rstudio	2021-05-24 7:48 PM	Application	6,574 KB

Fig. 1.23 Creating a shortcut for the RStudio application

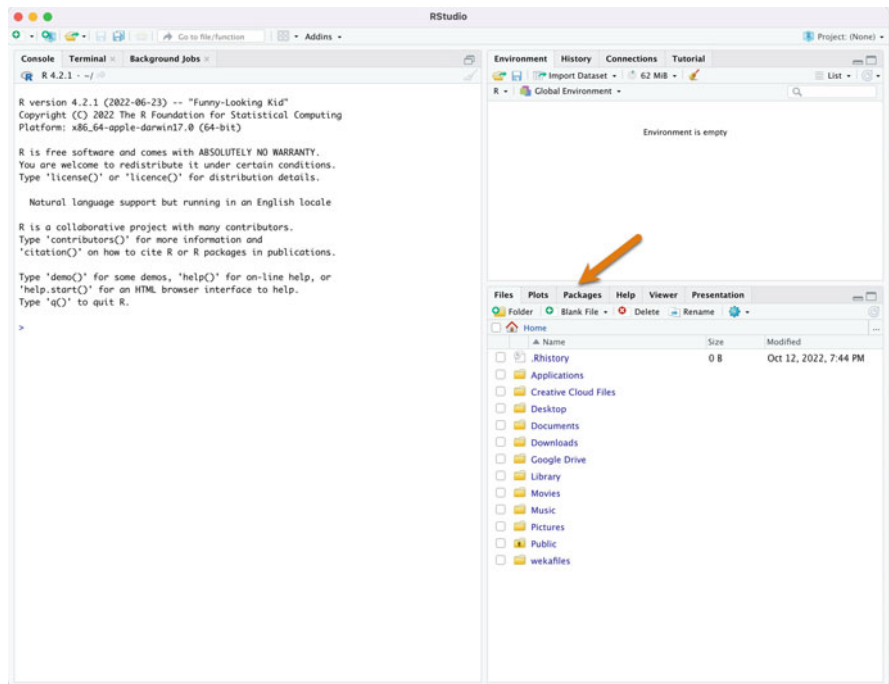


Fig. 1.24 Launching RStudio IDE

1.10.2.2 Install a Package

Packages are libraries that allow us to do specific tasks in RStudio (e.g., load a file, display a result, do an analysis). A package can be installed in the RStudio console using the following instructions:

1. Click on the Packages tab and click on the Install button (Fig. 1.25):
2. Install the readr package by typing “readr” and clicking on the Install button (Fig. 1.26).

1.10.2.3 Activate Package

1. To activate the readr library used to read csv and txt files, type “library(readr)” (Fig. 1.27):

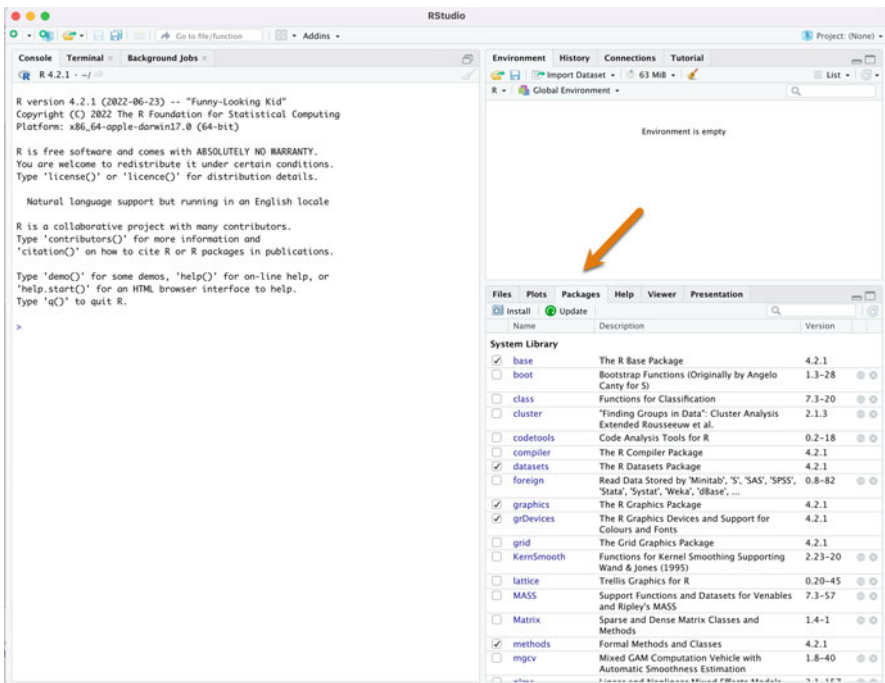


Fig. 1.25 Navigating RStudio Packages tab

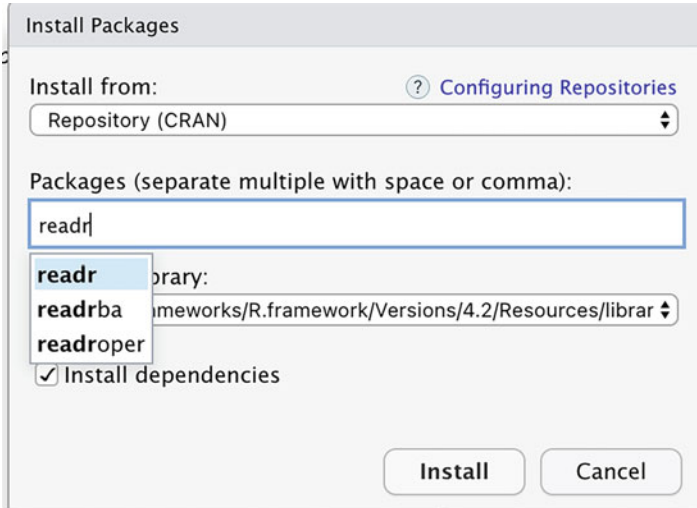


Fig. 1.26 Installing readr package for loading files

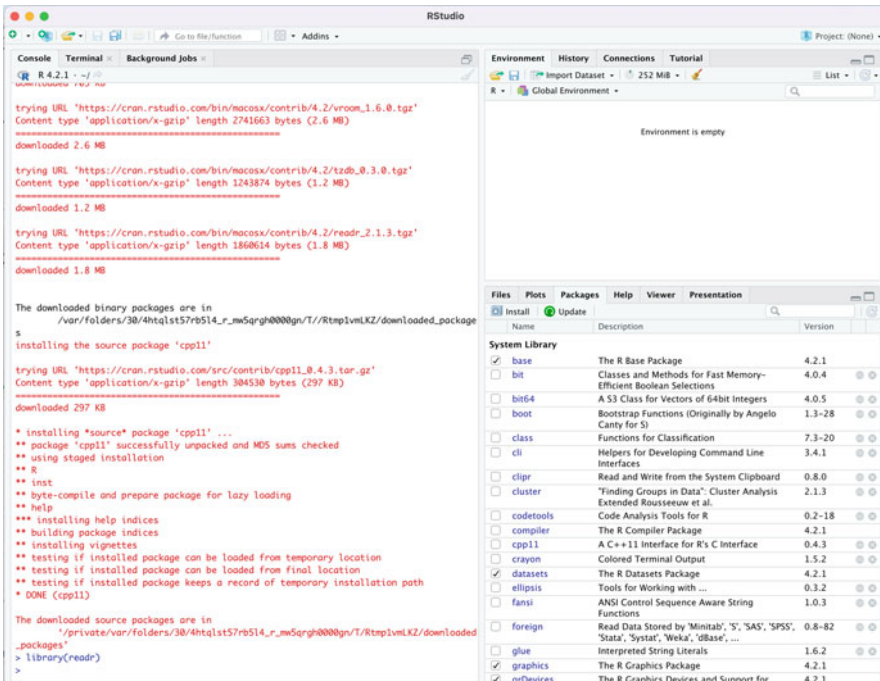


Fig. 1.27 Activating readr package for use

1.10.2.4 User Readr to Load Data

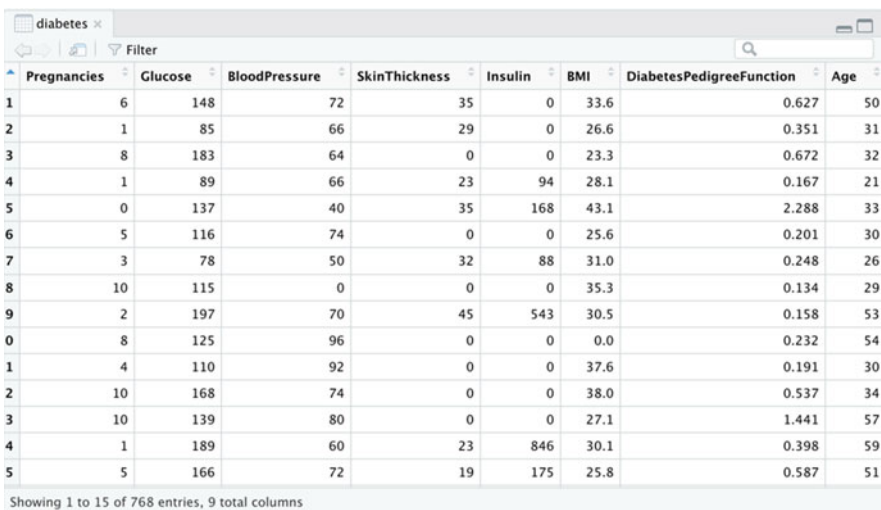
Different dataset types, such as txt, csv, and xlsx, can be imported into RStudio files (Fig. 1.28)

1. Download Diabetes.csv: Go to the following link <https://www.kaggle.com/uciml/pima-indians-diabetes-database/version/1> (or go to [kaggle.com](https://www.kaggle.com) and search for “Pima Indians Diabetes Database”).
2. Next, you will load the diabetes.csv file and plot it as a histogram. Under the main menu, click on the File menu/Import dataset/From text readr. In case you are prompted to download a library, accept to download it. Choose the input file from the folder where you have saved it and click Import.

1.10.2.5 Run a Function

The hist function can be used to visualize the data in a histogram. The hist function can present the data imported earlier (blood pressure vs. age) in a histogram (Fig. 1.29). Type the following:

```
hist(diabetes$BloodPressure,main="Blood Pressure Histogram",xlab =
"Blood Pressure",ylab = "Count",las=1) .
```



	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
1	6	148	72	35	0	33.6	0.627	50
2	1	85	66	29	0	26.6	0.351	31
3	8	183	64	0	0	23.3	0.672	32
4	1	89	66	23	94	28.1	0.167	21
5	0	137	40	35	168	43.1	2.288	33
6	5	116	74	0	0	25.6	0.201	30
7	3	78	50	32	88	31.0	0.248	26
8	10	115	0	0	0	35.3	0.134	29
9	2	197	70	45	543	30.5	0.158	53
0	8	125	96	0	0	0.0	0.232	54
1	4	110	92	0	0	37.6	0.191	30
2	10	168	74	0	0	38.0	0.537	34
3	10	139	80	0	0	27.1	1.441	57
4	1	189	60	23	846	30.1	0.398	59
5	5	166	72	19	175	25.8	0.587	51

Showing 1 to 15 of 768 entries, 9 total columns

Fig. 1.28 Importing dataset using the readr package

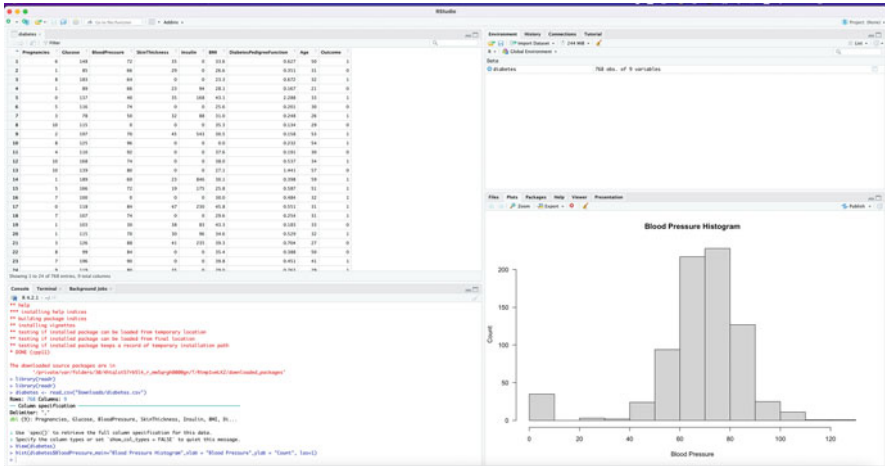


Fig. 1.29 Visualize data in a histogram

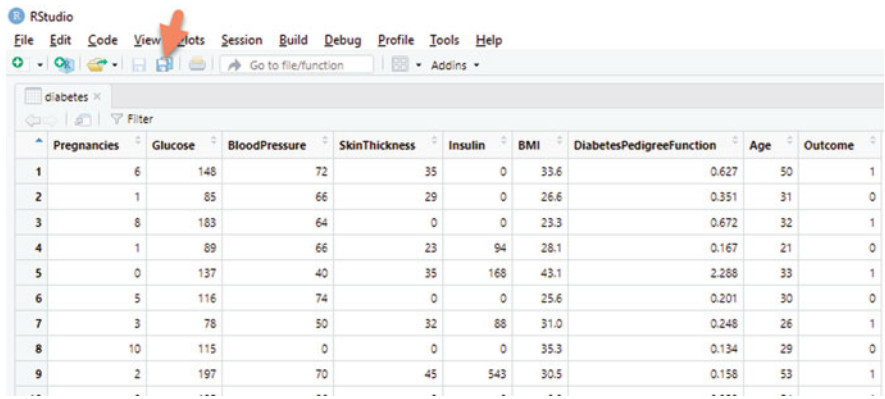


Fig. 1.30 Save progress status in RStudio

1.10.2.6 Save Status

To save your progress, use the ctrl + S shortcut or click the blue Save button in the main menu (Fig. 1.30):

1.10.3 Introduction to Python and Jupyter Notebook IDE

Python is an interpreted programming language. It is characterized by human readability; it is an important programming language in machine learning and artificial intelligence due to its flexible libraries and ease of use.

In this book, Jupyter Notebook IDE is used for Python labs and examples.

1.10.3.1 Python Download and Installation

Python programming language can be installed from the following location: [Python.org](https://python.org). Below are instructions for python download and installation for Python v3.9.6.

1. Download the Windows installer to your computer location (Fig. 1.31).
2. Install Python v3.9.6 (64-bit) using the “Customize installation” option; please follow the screenshots carefully by checking the right checkboxes as indicated (Figs. 1.32, 1.33, and 1.34).
3. To validate the installation, open a terminal (open “cmd” in Windows; open a terminal in macOS) and type “*python -v*.”

Version	Operating System	Description	MD5 Sum	File Size	GPG	Sigstore
Gzipped source tarball	Source release		fbc3ff11893916ad1756b15c8a48834	26015299	SIG CRT SIG	
XZ compressed source tarball	Source release		e92356b012ed4d0e09675131d39b1bde	19619508	SIG CRT SIG	
macOS 64-bit universal2 installer	macOS	for macOS 10.9 and later	eb11b5816b1a37d934070145391eadfe	40883084	SIG CRT SIG	
Windows embeddable package (32-bit)	Windows		e0dbee095e5963b26b8bf258fd2b9f41	7617241	SIG CRT SIG	
Windows embeddable package (64-bit)	Windows		923be16c4cef2474b7982d16cea60ddb	8592015	SIG CRT SIG	
Windows help file	Windows		0cbb41f049c8f496f4fb18d84430d9a	9379210	SIG CRT SIG	
Windows installer (32-bit)	Windows		10efcd9a8777fe84f9a9c583d074e632	27820784	SIG CRT SIG	
Windows installer (64-bit)	Windows	Recommended	308a3d095311fbc82e5c696ab4036251	28978512	SIG CRT SIG	

Fig. 1.31 Install Python v3.9.6 Windows installer



Fig. 1.32 Setting up Python 3.9.6

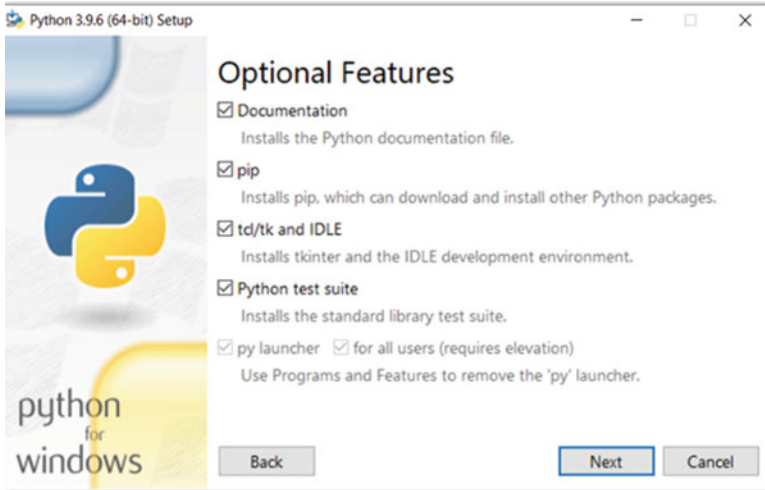


Fig. 1.33 Configuring Python 3.9.6 features

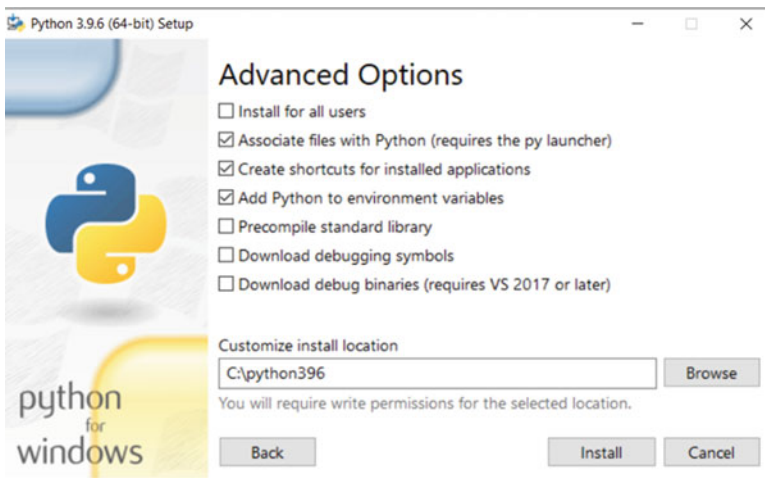


Fig. 1.34 Configuring more Python v3.9.6 installation options

1.10.3.2 Jupyter Download and Installation

1. To install “Jupyter Notebook” IDE, open a terminal and type “*pip install jupyter*”; if the command does not work then type “*pip3 install jupyter*” (Fig. 1.35).

```

Administrator: Command Prompt

C:\Windows\System32>pip install jupyter
Collecting jupyter
  Using cached jupyter-1.0.0-py2.py3-none-any.whl (2.7 kB)
Requirement already satisfied: ipykernel in c:\users\elmorr\appdata\roaming\python\python310\site-packages (from jupyter) (6.5.1)
Collecting qtconsole
  Downloading qtconsole-5.3.2-py3-none-any.whl (120 kB)
----- 120.0/120.0 kB 1.0 MB/s eta 0:00:00
Requirement already satisfied: ipywidgets in c:\users\elmorr\appdata\roaming\python\python310\site-packages (from jupyter) (7.6.5)
Collecting jupyter-console
  Downloading jupyter_console-6.4.4-py3-none-any.whl (22 kB)
Collecting nbconvert
  Downloading nbconvert-7.2.1-py3-none-any.whl (271 kB)
----- 271.0/271.0 kB 8.4 MB/s eta 0:00:00
Collecting notebook
  Downloading notebook-6.4.12-py3-none-any.whl (9.9 MB)
----- 9.9/9.9 MB 63.6 MB/s eta 0:00:00
Requirement already satisfied: ipython>=7.23.1 in c:\users\elmorr\appdata\roaming\python\python310\site-packages (from ipykernel->jupyter) (7.30.0)
Requirement already satisfied: traitlets<6.0,>=5.1.0 in c:\users\elmorr\appdata\roaming\python\python310\site-packages (from ipykernel->jupyter) (5.1.1)
Requirement already satisfied: tornado<7.0,>=4.2 in c:\users\elmorr\appdata\roaming\python\python310\site-packages (from ipykernel->jupyter) (6.1)
Requirement already satisfied: matplotlib-inline<0.2.0,>=0.1.0 in c:\users\elmorr\appdata\roaming\python\python310\site-packages (from ipykernel->jupyter) (0.1.3)
Collecting jupyter-client<8.0
  Downloading jupyter_client-7.4.2-py3-none-any.whl (132 kB)
----- 132.2/132.2 kB 0.1 MB/s eta 0:00:00
Requirement already satisfied: debugpy<2.0,>=1.0.0 in c:\users\elmorr\appdata\roaming\python\python310\site-packages (from ipykernel->jupyter) (1.5.1)
Requirement already satisfied: ipython-genutils<=0.2.0 in c:\users\elmorr\appdata\roaming\python\python310\site-packages (from ipywidgets->jupyter) (0.2.0)
Requirement already satisfied: widgetsnbextension<=3.5.0 in c:\users\elmorr\appdata\roaming\python\python310\site-packages (from ipywidgets->jupyter) (3.5.2)
Collecting nbformat<=4.2.0
  Downloading nbformat-5.7.0-py3-none-any.whl (77 kB)
----- 77.1/77.1 kB 4.2 MB/s eta 0:00:00
Requirement already satisfied: jupyterlab-widgets<=1.0.0 in c:\users\elmorr\appdata\roaming\python\python310\site-packages (from ipywidgets->jupyter) (1.0.2)
Requirement already satisfied: prompt-toolkit!=3.0.0,!<3.0.1,<3.1.0,>=2.0.0 in c:\users\elmorr\appdata\roaming\python\python310\site-packages (from jupyter-console->jupyter) (3.0.23)
Requirement already satisfied: pygments in c:\users\elmorr\appdata\roaming\python\python310\site-packages (from jupyter-console->jupyter) (2.10.0)
Collecting jupyter-core<=4.7
  Downloading jupyter_core-4.11.1-py3-none-any.whl (88 kB)

```

Fig. 1.35 Installing Jupyter Notebook IDE

2. We need a useful Python package called “pandas” to manipulate data. In order to use pandas, we need first to install them. To do so we will use the command `pip install`.

On Windows, open a terminal (cmd) *as an administrator*; you can do so by right clicking on cmd and choosing Run As Administrator. In the terminal, type “`pip install pandas`” (Fig. 1.36).

On macOS, open the terminal and write “`sudo pip install pandas`.”

macOS will ask you for your password; the system assumes you have administrative powers. Enter the password and the library package will be installed (Fig. 1.37).

3. Using the same strategy, install `matplotlib` and `openpyxl` libraries (Fig. 1.38):

```

pip install matplotlib
pip install openpyxl

```

For MacOS, launch Jupyter Notebook by typing the following in the terminal: “Jupyter Notebook”. For windows type “`python -m jupyter notebook`”. Then, click “New” button and choose “Python 3” to create a notebook (Figs. 1.39 and 1.40).

```

Administrator: Command Prompt
Microsoft Windows [Version 10.0.22000.100]
(c) Microsoft Corporation. All rights reserved.

C:\WINDOWS\system32>pip install pandas
Collecting pandas
  Using cached pandas-1.3.1-cp39-cp39-win_amd64.whl (10.4 MB)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\python39\lib\site-packages (from pandas) (1.21.1)
Requirement already satisfied: numpy>=1.17.3 in c:\python39\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2017.3 in c:\python39\lib\site-packages (from pandas) (2021.1)
Requirement already satisfied: six>=1.5 in c:\python39\lib\site-packages (from python-dateutil>=2.7.3->pandas) (1.16.0)
Installing collected packages: pandas
Successfully installed pandas-1.3.1

C:\WINDOWS\system32>

```

Fig. 1.36 Installing pandas package to work with data frames for Windows users

```

MyMacBook2019:~ elmorr$ sudo pip install pandas
Password:
WARNING: The directory '/Users/elmorr/Library/Caches/pip' or its parent directory
y is not owned or is not writable by the current user. The cache has been disabl
ed. Check the permissions and owner of that directory. If executing pip with sud
o, you should use sudo's -H flag.
Collecting pandas
  Downloading pandas-1.3.1-cp39-cp39-macosx_10_9_x86_64.whl (11.3 MB)
    |-----| 11.3 MB 201 kB/s
Requirement already satisfied: python-dateutil>=2.7.3 in /Library/Frameworks/Pyt
hon.framework/Versions/3.9/lib/python3.9/site-packages (from pandas) (2.8.2)
Collecting pytz>=2017.3
  Downloading pytz-2021.1-py2.py3-none-any.whl (510 kB)
    |-----| 510 kB 9.2 MB/s
Collecting numpy>=1.17.3
  Downloading numpy-1.21.1-cp39-cp39-macosx_10_9_x86_64.whl (17.0 MB)
    |-----| 17.0 MB 2.7 MB/s
Requirement already satisfied: six>=1.5 in /Library/Frameworks/Python.framework/

```

Fig. 1.37 Installing pandas package to work with data frames for Mac users

```

MyMacBook2019:~ elmorr$ sudo pip install pandas
Password:
WARNING: The directory '/Users/elmorr/Library/Caches/pip' or its parent directory
y is not owned or is not writable by the current user. The cache has been disabl
ed. Check the permissions and owner of that directory. If executing pip with sud
o, you should use sudo's -H flag.
Collecting pandas
  Downloading pandas-1.3.1-cp39-cp39-macosx_10_9_x86_64.whl (11.3 MB)
    |-----| 11.3 MB 201 kB/s
Requirement already satisfied: python-dateutil>=2.7.3 in /Library/Frameworks/Pyt
hon.framework/Versions/3.9/lib/python3.9/site-packages (from pandas) (2.8.2)
Collecting pytz>=2017.3
  Downloading pytz-2021.1-py2.py3-none-any.whl (510 kB)
    |-----| 510 kB 9.2 MB/s
Collecting numpy>=1.17.3
  Downloading numpy-1.21.1-cp39-cp39-macosx_10_9_x86_64.whl (17.0 MB)
    |-----| 17.0 MB 2.7 MB/s
Requirement already satisfied: six>=1.5 in /Library/Frameworks/Python.framework/

```

Fig. 1.38 Installing matplotlib and openpyxl packages

```
C:\Users\Administrator>jupyter notebook
[I 17:43:23.779 NotebookApp] Serving notebooks from local directory: C:\Users\Administrator
[I 17:43:23.779 NotebookApp] Jupyter Notebook 6.4.0 is running at:
[I 17:43:23.779 NotebookApp] http://localhost:8888/?token=4a894990301ccad2a5a8c79b6957c464bdb3ddcddd98abc
[I 17:43:23.779 NotebookApp] or http://127.0.0.1:8888/?token=4a894990301ccad2a5a8c79b6957c464bdb3ddcddd98abc
[I 17:43:23.779 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 17:43:23.807 NotebookApp]

To access the notebook, open this file in a browser:
File:///C:/Users/Administrator/AppData/Roaming/jupyter/runtime/nbserver-11988-open.html
Or copy and paste one of these URLs:
http://localhost:8888/?token=4a894990301ccad2a5a8c79b6957c464bdb3ddcddd98abc
or http://127.0.0.1:8888/?token=4a894990301ccad2a5a8c79b6957c464bdb3ddcddd98abc
```

Fig. 1.39 Launching Jupyter Notebook IDE



Fig. 1.40 Notebook webpage

1.10.3.3 Load Data and Plot It Visually

Python code can be added in the Jupyter Notebook IDE file, and every line can be executed using the “Run” button for every line. It is important to note that code on all lines can be run once by doing the following: Under the Cell menu, click “Run All.” This is shown below in Fig. 1.41.

The code below allows us to read the blood pressure measurement from the diabetes.xlsx file and plot it in a histogram. Open diabetes.csv and save it as diabetes.xlsx, then follow the instructions below (Fig. 1.42):

```
Type: import pandas as pd
Type: import matplotlib.pyplot as plt
Type: df=pd.read_excel("diabetes.xlsx")
Type: dfplt=df.plot(kind="hist")
```

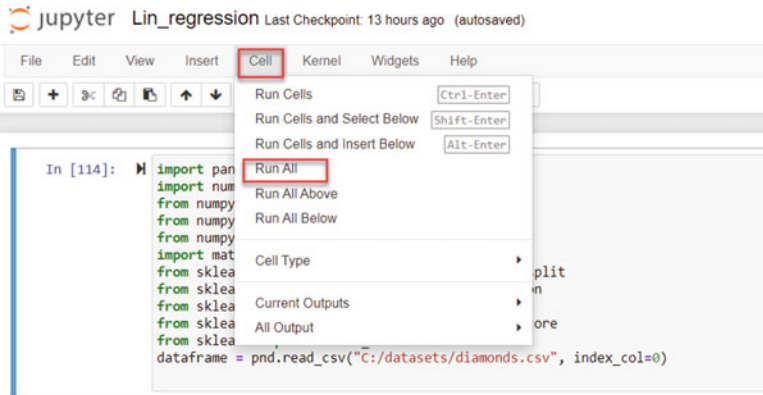


Fig. 1.41 Load data in Jupyter Notebook and execute code

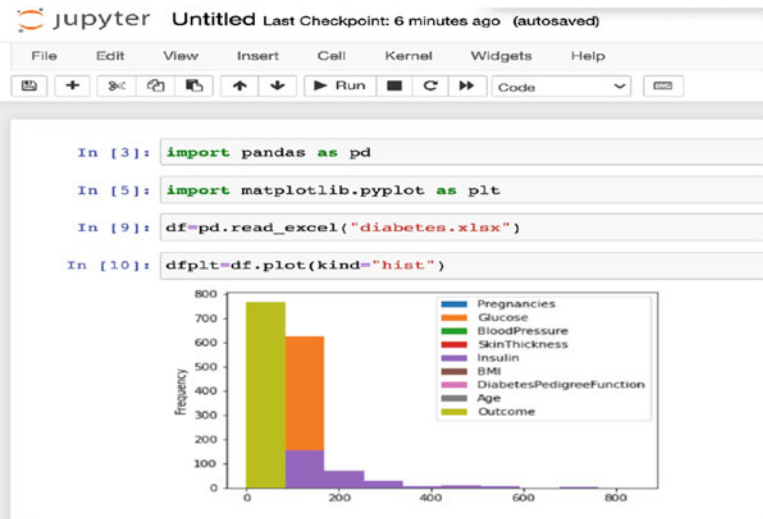


Fig. 1.42 Visualizing blood pressure measurements in a graph

1.10.3.4 Save the Execution

All that we have done can be saved in a file with the extension “ipynb.” This file can be loaded later in Jupyter Notebook IDE for any updates or changes to continue working from where you left. Click on “File,” then choose “Save” or click on the save icon.

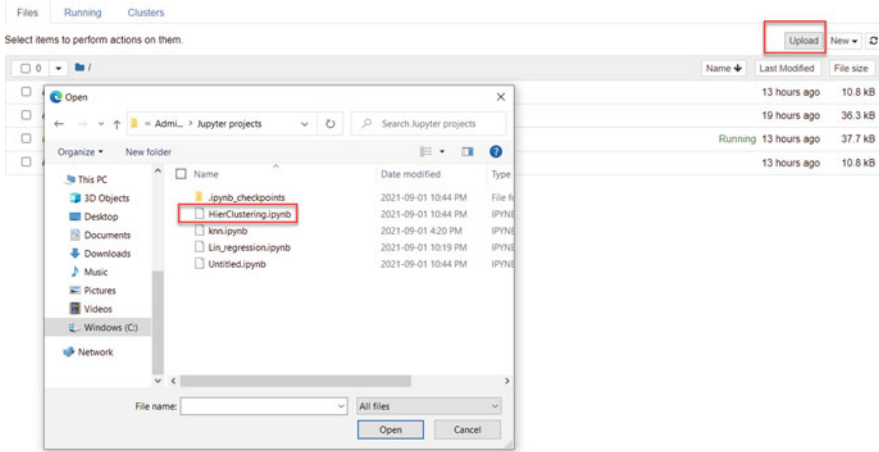


Fig. 1.43 Upload Jupyter Notebook files into the application

1.10.3.5 Load a Saved Execution

Go to your workspace folder in the command line and lunch Jupyter Notebook. Then, double-click on the file that you need to continue working on.

1.10.3.6 Upload a Jupyter Notebook File

You can also upload files into Jupyter Notebook through the application interface. After launching the program, click on the Upload button and upload all the files that you want to upload, as shown in Fig. 1.43.

1.10.4 Do It Yourself

The following problem is to try by yourself:

1. Weka is a machine learning software. Install Weka from the following link: https://waikato.github.io/weka-wiki/downloading_weka/
2. Jupyter lab is the next generation Jupyter Notebook interface. Install Jupyter lab from <https://jupyter.org/>. We suggest that you use Jupyter lab instead of Jupyter Notebook.
3. Download any dataset from Find Open Datasets and Machine Learning Projects | Kaggle and plot the data visually in R.
4. Download any dataset from Find Open Datasets and Machine Learning Projects | Kaggle and plot the data visually in Jupyter Notebook or Jupyter lab.
5. Try Colabroatory, also known as Colab, the online Python development environment provided by Google: <https://colab.research.google.com/>. We strongly advise you to use either Colab or JupyterLab for your projects.

References

1. E. Alpaydin, *Machine Learning: The New AI* (MIT Press, 2016)
2. H. Hassani, P. Amiri Andi, A. Ghodsi, K. Norouzi, N. Komendantova, S. Unger, Shaping the future of smart dentistry: From Artificial Intelligence (AI) to Intelligence Augmentation (IA). *IoT* **2**(3), 510–523 (2021)
3. R.E. Neapolitan, X. Jiang, *Artificial Intelligence: With an Introduction to Machine Learning* (CRC Press, 2018)
4. A.L. Fradkov, Early history of machine learning. *IFAC-PapersOnLine* **53**(2), 1385–1390 (2020)
5. J.A. Nichols, H.W.H. Chan, M.A. Baker, Machine learning: Applications of artificial intelligence to imaging and diagnosis. *Biophys. Rev.* **11**(1), 111–118 (2019)
6. Q. Bi, K.E. Goodman, J. Kaminsky, J. Lessler, What is machine learning? A primer for the epidemiologist. *Am. J. Epidemiol.* **188**(12), 2222–2239 (2019)
7. X. Zhu, A.B. Goldberg, Introduction to semi-supervised learning. *Synth. Lect. Artif. Intell. Mach. Learn.* **3**(1), 1–130 (2009)
8. I.H. Sarker, Machine learning: Algorithms, real-world applications and research directions. *SN Comput. Sci.* **2**(3), 1–21 (2021)
9. N. Kalé, N. Jones, *Practical Analytics* (Epistemy Press, 2015)
10. R. Sharda, D. Delen, E. Turban, *Business Intelligence: A Managerial Perspective on Analytics: A Managerial Perspective on Analytics* (Pearson, 2015), pp. 416–416
11. K.C. Laudon, J.P. Laudon, *Management Information Systems: Managing the Digital Firm* (Pearson, 2017)
12. K.E.S.C.S. Pearlson, D.F. Galletta, *Managing and Using Information Systems a Strategic Approach* (John Wiley & Sons, 2016)
13. C. El Morr, H. Ali-Hassan, *Analytics in Healthcare: A Practical Introduction* (Springer, 2019)
14. P. Drucker, The coming of the new organization. *Harv. Bus. Rev.* **Jan–Feb**, 45–53 (1988)
15. T. Davenport, *Information Ecology* (Oxford University Press, New York, 1997)
16. SAS, Big Data - What it is and why it matters. https://www.sas.com/en_ca/insights/big-data/what-is-big-data.html. Accessed
17. D. Faggella, Where Healthcare’s big data actually comes from. 11 Jan 2018. [Online]. Available: <https://www.techemergence.com/where-healthcares-big-data-actually-comes-from/>
18. D. Pogue, Exclusive: Fitbit’s 150 billion hours of heart data reveal secrets about health. August 27, 2018. [Online]. Available: <https://finance.yahoo.com/news/exclusive-fitbits-150-billion-hours-heart-data-reveals-secrets-human-health-133124215.html?linkId=56096180>
19. J. Bresnick, Understanding the Many V’s of Healthcare Big data analytics. 5 June 2017. [Online]. Available: <https://healthitanalytics.com/news/understanding-the-many-vs-of-healthcare-big-data-analytics>
20. R. Sharda, D. Delen, E. Turban, *Business Intelligence: A Managerial Perspective on Analytics* (Prentice Hall Press, 2015)
21. T. Economist, *Data Is Giving Rise to a New Economy* (The Economist, 6 May 2017)
22. R. Sharda, D. Delen, E. Turban, J. Aronson, T.P. Liang, *Business Intelligence and Analytics: Systems for Decision Support* (Prentice Hall Press, 2014)
23. K.C. Laudon, J.P. Laudon, *Essentials of Management Information Systems* (Pearson Upper Saddle River, 2011)
24. C. Ballard, D.M. Farrell, A. Gupta, C. Mazuela, S. Vohnik, *Dimensional Modeling: In a Business Intelligence Environment* (IBM Redbooks, 2012)
25. K.E. Pearlson, C.S. Saunders, D.F. Galletta, *Managing and Using Information Systems a Strategic Approach* (John Wiley & Sons, 2016)
26. R. Sharda, D. Delen, E. Turban, *Business Intelligence: A Managerial Perspective on Analytics* (Prentice Hall Press, 2013)
27. H. Mailvaganam, Introduction to OLAP. http://www.dwreview.com/OLAP/Introduction_OLAP.html. Accessed

28. R. Sharda, D. Delen, E. Turban, *Business Intelligence: A managerial Perspective on Analytics* (Prentice Hall Press, 2014)
29. L. Madsen, Business intelligence an introduction, in *Healthcare Business Intelligence: A Guide to Empowering Successful Data Reporting and Analytics*, (Wiley, 2012)
30. K.D. Lawrence, R. Klimberg, *Contemporary Perspectives in Data Mining*, vol 1 (Information Age Publishing, 2013)
31. M. K. Pratt, Business intelligence vs. business analytics: Where BI fits into your data strategy. *CIO Magazine*, 2017. Available: <https://www.cio.com/article/2448992/business-intelligence/business-intelligence-vs-business-analytics-where-bi-fits-into-your-data-strategy.html>
32. Rose Business Technologies, *Descriptive Diagnostic Predictive Prescriptive Analytics*. Rose Business Technologies. <http://www.rosebt.com/blog/descriptive-diagnostic-predictive-prescriptive-analytics>. Accessed 26 April 2018
33. J. Bresnick, Healthcare big data analytics: from description to prescription. <https://healthitanalytics.com/news/healthcare-big-data-analytics-from-description-to-prescription>. Accessed
34. S. Maloney, Making Sense of Analytics. Presented at the eHealth2018, Toronto ON. [Online]. Available: <http://www.healthcareimc.com/main/making-sense-of-analytics/>
35. R.S. Uberoi, U. Gupta, A. Sibal, Root cause analysis in healthcare. *Apollo Med.* **1**(1), 60–63 (9 Jan 2004). [https://doi.org/10.1016/S0976-0016\(12\)60044-1](https://doi.org/10.1016/S0976-0016(12)60044-1)
36. W.E. Fassett, Key performance outcomes of patient safety curricula: root cause analysis, failure mode and effects analysis, and structured communications skills. *Am. J. Pharm. Educ.* **75**(8), 164 (10 Oct 2011). <https://doi.org/10.5688/ajpe758164>
37. R. Ursprung, J. Gray, Random safety auditing, root cause analysis, failure mode and effects analysis. *Clin. Perinatol.* **37**(1), 141–165 (Mar 2010). <https://doi.org/10.1016/j.clp.2010.01.008>
38. M. Liao, Y. Li, F. Kianifard, E. Obi, S. Arcona, Cluster analysis and its application to healthcare claims data: A study of end-stage renal disease patients who initiated hemodialysis. *BMC Nephrol.* **17**, 25 (2016). <https://doi.org/10.1186/s12882-016-0238-2>
39. M. Chowdhury, A. Apon, K. Dey, *Data Analytics for Intelligent Transportation Systems* (Elsevier Science, 2017)
40. H.H. Hijazi, H.L. Harvey, M.S. Alyahya, H.A. Alshraideh, R.M. Al Abdi, S.K. Parahoo, The impact of applying quality management practices on patient centeredness in jordanian public hospitals: results of predictive modeling. *Inquiry: J Medical Care Organization, Provision and Financing* **55**, 46958018754739 (Jan–Dec 2018). <https://doi.org/10.1177/0046958018754739>
41. F. Noviyanti, Y. Hosotani, S. Koseki, Y. Inatsu, S. Kawasaki, Predictive modeling for the growth of salmonella enteritidis in chicken juice by real-time polymerase chain reaction. *Foodborne Pathog. Dis.* **15**(7), 406–412 (2 Apr 2018). <https://doi.org/10.1089/fpd.2017.2392>
42. M.M. Safaee et al., Predictive modeling of length of hospital stay following adult spinal deformity correction: Analysis of 653 patients with an accuracy of 75% within 2 days. *World Neurosurg* **115**, e422–e427 (17 Apr 2018). <https://doi.org/10.1016/j.wneu.2018.04.064>
43. B. Baessler, M. Mannil, D. Maintz, H. Alkadhi, R. Manka, Texture analysis and machine learning of non-contrast T1-weighted MR images in patients with hypertrophic cardiomyopathy-preliminary results. *Eur. J. Radiol.* **102**, 61–67 (May 2018). <https://doi.org/10.1016/j.ejrad.2018.03.013>
44. P. Karisani, Z.S. Qin, E. Agichtein, Probabilistic and machine learning-based retrieval approaches for biomedical dataset retrieval. *Database: J. Biol. Databases Curation* **2018**, bax104 (1 Jan 2018). <https://doi.org/10.1093/database/bax104>
45. M.R. Schadler, A. Warzybok, B. Kollmeier, Objective prediction of hearing aid benefit across listener groups using machine learning: Speech recognition performance with binaural noise-reduction algorithms. *Trends Hear.* **22**, 2331216518768954 (Jan–Dec 2018). <https://doi.org/10.1177/2331216518768954>
46. Y. Wu, K. Doi, C.E. Metz, N. Asada, M.L. Giger, Simulation studies of data classification by artificial neural networks: Potential applications in medical imaging and decision making. *J. Digit. Imaging* **6**(2), 117–125 (May 1993)

47. J. Zhang, M. Liu, D. Shen, Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks. *IEEE Trans. Image Process.* **26**(10), 4753–4764 (Oct 2017). <https://doi.org/10.1109/tip.2017.2721106>
48. E. Chalmers, D. Hill, V. Zhao, E. Lou, Prescriptive analytics applied to brace treatment for AIS: A pilot demonstration. *Scoliosis* **10**(Suppl 2), S13 (2015). <https://doi.org/10.1186/1748-7161-10-s2-s13>
49. F. Devriendt, D. Moldovan, W. Verbeke, A literature survey and experimental evaluation of the state-of-the-art in uplift modeling: A stepping stone toward the development of prescriptive analytics. *Big Data* **6**(1), 13–41 (Mar 2018). <https://doi.org/10.1089/big.2017.0104>
50. S. Van Poucke, M. Thomeer, J. Heath, M. Vukicevic, Are Randomized controlled trials the (G)old standard? From clinical intelligence to prescriptive analytics. *Journal of medical Internet Research* **18**(7), e185 (6 Jul 2016). <https://doi.org/10.2196/jmir.5549>
51. G.K. Alexander, S.B. Canclini, J. Fripp, W. Fripp, Waterborne disease case investigation: Public health nursing simulation. *J. Nurs. Educ.* **56**(1), 39–42 (1 Jan 2017). <https://doi.org/10.3928/01484834-20161219-08>
52. M. Lee, Y. Chun, D.A. Griffith, Error propagation in spatial modeling of public health data: A simulation approach using pediatric blood lead level data for Syracuse, New York. *Environ. Geochem. Health* **40**(2), 667–681 (Apr 2018). <https://doi.org/10.1007/s10653-017-0014-7>
53. M. Moessner, S. Bauer, Maximizing the public health impact of eating disorder services: A simulation study. *Int. J. Eat. Disord.* **50**(12), 1378–1384 (Dec 2017). <https://doi.org/10.1002/eat.22792>
54. O. El-Rifai, T. Garaix, V. Augusto, X. Xie, A stochastic optimization model for shift scheduling in emergency departments. *Health Care Manag. Sci.* **18**(3), 289–302 (Sep 2015). <https://doi.org/10.1007/s10729-014-9300-4>
55. A. Jeremic, E. Khoshrowshahli, Detecting breast cancer using microwave imaging and stochastic optimization. Conference proceedings: . . . Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference **2015**, 89–92 (2015). <https://doi.org/10.1109/embc.2015.7318307>
56. A. Legrain, M.A. Fortin, N. Lahrichi, L.M. Rousseau, Online stochastic optimization of radiotherapy patient scheduling. *Health Care Manag. Sci.* **18**(2), 110–123 (Jun 2015). <https://doi.org/10.1007/s10729-014-9270-6>
57. M.A. Christodoulou, C. Kontogeorgou, Collision avoidance in commercial aircraft free flight via neural networks and non-linear programming. *Int. J. Neural Syst.* **18**(5), 371–387 (Oct 2008). <https://doi.org/10.1142/s0129065708001658>
58. S.I. Saffer, C.E. Mize, U.N. Bhat, S.A. Szygenda, Use of non-linear programming and stochastic modeling in the medical evaluation of normal-abnormal liver function. *I.E.E.E. Trans. Biomed. Eng.* **23**(3), 200–207 (May 1976)
59. G.H. Simmons, J.M. Christenson, J.G. Kereiakes, G.K. Bahr, A non-linear programming method for optimizing parallel-hole collimator design. *Phys. Med. Biol.* **20**(3), 771–788 (Sep 1975)
60. I. Podolak, Making sense of analytics. Presented at the eHealth 2017, Toronto ON, 2017. [Online]. Available: <http://www.healthcareimc.com/main/making-sense-of-analytics/>