|  |
| --- |
| Data Analysis  Study and Findings  15-02-2023 |

|  |
| --- |
| Parisa Azimivahdat |

|  |  |
| --- | --- |
| Parisa Azimivahdat, Student  parisaAzimivahdat.bookcamp@justit.co.uk |  |



CONTETS

Introduction3

CHAPTER 14

Basic Concepts

Data Structure

Data Types

Big Data

Sorting Techniques

CHAPTER 212

IT Laws and Regulations

The CIA Triad

Related Legislation

CHAPTER 314

Excel

CHAPTER 418

Power Query

CHAPTER 521

Power BI

CHAPTER 623

Tableau

CHAPTER 725

SQL Problem-solving

|  |
| --- |
| Introduction A report will cover all learning on the Data Skills Bootcamp about understanding the concepts and work of a Data Analyst. After presenting a summary of combined lecture notes from the educational share point, chosen tasks are performed and presented with screen shots and explanations to test functionality of the related software. |
| Why Every Business Student Needs Data Analysis Skills |
| The following software were tested and used to produce the tasks.   * Excel   Creating a searchable list on the ‘Back End’ and ‘Front End’ worksheets. A list of names is made to search for the letter entered a search box. On the front end the search box can be extended to display the entire record for that name which can be chosen from a drop down list. |  |  |

|  |  |
| --- | --- |
| Data analytics là gì? Cách trở thành một data analyst? |  |
| *Power Query*  Data, “NewJerseySales”, is sourced in the excel folder posted on JustIt’s share point, downloaded, and saved on the home PC.  Step 1: Get Data  Step 2: Transform Data, replace null values and round up Decimal numbers to currency.  Step 3: Create a pivot table and bar chart. Style the bar chart.   * *Power BI*   “NewJerseySales” is used for data. The data has been cleaned in Power Query. The following components are created to make up a dashboard presenting the sales in New Jersey in 2021: New columns, box with the heading, calculation boxes showing the sum of a single column, line chart, area chart, pie chart and funnel chart and also a zoom bar.   * *Tableau*   “NewJerseySales” is used for this dashboard in Tableau. Three worksheets are used to make the dashboard.  Worksheet 1: The map  Worksheet 2: The Horizontal Bar Chart  Worksheet 3: Packed Bubbles |
| CHAPTER 1 A picture containing background pattern  Description automatically generated  What is data and how much of it can we find out there? |  |
| Basic ConceptsAll manual and automated inputs into the interface of a computer are considered data. Words, audio, numbers, images, and measurements are what humans enter and download using input devices such as a keyboard, mouse, touchpad, or screen touch. Available data can be enhanced for usage, making it possible to modify and transport it. Raw data or primary data is the state of untreated data generated globally every day by more than 3.7 billion users.What does Data look like? A computer processes in the binary number system, consisting of 1s and 0s, the 1 is ON and 0 is OFF state of the electrical current. Users enter data by using interactive entry- or search boxes which then translate into binary numbers by the ASCII number table. The ASCII table is the key covering columns for decimal numbers, hexadecimal numbers, and char which stands for characters including letters in upper and lower case and symbols. |

## Data Structures

There are basic concepts referring to how data is generally organized, such as records and fields in a file. Another is grouped items which groups a person’s first and last names in separate columns. These columns together identify a person and can be used to create a personal ID, number, or email address for them. Among these concepts are also entities which stand for the class of objects with the same attribute or property. An important source of primitive data is a Data Lake which contains all types of data in a range from structured to nearly untransformed state without data cleaning or packaging. (ref to data & info slide 20)

## What do Data structures do?

Data Structures fit a data model as part of the design of the software. The data structure controls the flow of data inputs which makes it easier to process, visualize and transport the data. Data is structured by its properties and sorted into categories. This way calculations, comparisons, and evaluation of the data can be done. Structured data can be used with analyzing tools for interpretation of its meaning and potential so it can be turned into intelligent information. Business companies extract information from their data to study and engage the public and meet their business targets.

## Business Intelligence

Raw data goes through four processing stages in data analysis before gaining its full potential. These stages are Data, Information, Knowledge, and Wisdom.

Business intelligence can be gained by applying three different kinds of data analysis. The first is the descriptive analysis which deals with historical data. The Second type is the predictive analysis which is establishing a trend in the past and current data sets in order to predict a future trend. Predictive analysis is using simulation models and forecasting to give estimations. Thirdly, we have prescriptive analysis which recommends winning decisions and the next intelligent steps to be taken in order to achieve the business’s set of targets.

Qualitative: describes properties, in text form

Nominal: unordered categories

Ordinal: ordered categories

Categorical data: characteristics expressed in text, can be numbers but without mathematical meaning

Quantitative: measurements, values, or counts on a mathematical scale, in number form

Interval: values ordered on a scale with negative numbers and zero

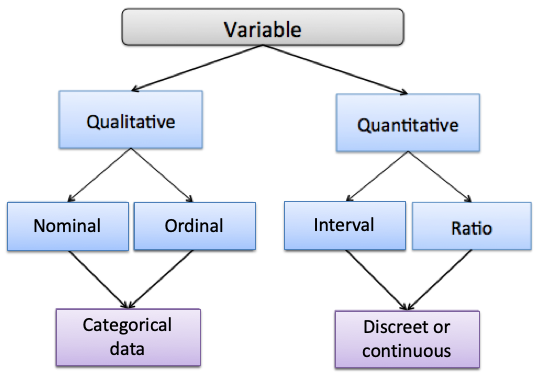
Ratio: ordered values on a scale with zero denoting the absence of something

Discrete: count data

continuous: continuous measurements

## Data Types

There are different data types for different users. Compare an engineering program, a word processing software, and an audio recording application for the type of data they process.



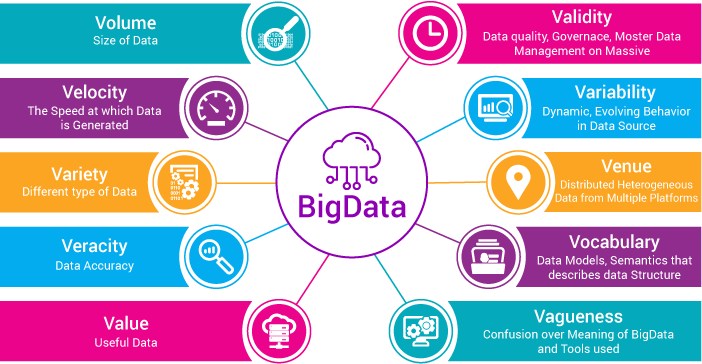
[See here](https://bookdown.org/curleyjp0/psy317l_guides5/introduction-to-data-carpentry.html).

## A picture containing text Description automatically generated

Most common data types with size and range. Signed integers contain negative numbers.

## Big Data

Data stretching over vast and various areas of human interaction and behaviour are called Big Data for their complexity, growth rate, and size. It can be in structured, unstructured, or semi-structured form. We need advanced tools when analysing big data, for example, macro tools for recognising patterns to be able to forecast and predict. An example of this is the weather forecast which has been an increasingly more precise science with the advance of Macro analysing tools. Big Data can be open-sourced and free for the public to use and distribute, or it can be private as access is regulated by the law. According to Paul Zikopaulos, Vice President of IBM technologies, it takes a minimum of 200-600 TB for a data bulk to qualify as Big.

Data.  [](https://towardsdatascience.com/big-data-analysis-spark-and-hadoop-a11ba591c057)

[There are four characteristics of Big Data:](https://towardsdatascience.com/big-data-analysis-spark-and-hadoop-a11ba591c057)

* [Volume: its size](https://towardsdatascience.com/big-data-analysis-spark-and-hadoop-a11ba591c057)
* [Velocity: its rate of generation](https://towardsdatascience.com/big-data-analysis-spark-and-hadoop-a11ba591c057)
* [Variety: its data types](https://towardsdatascience.com/big-data-analysis-spark-and-hadoop-a11ba591c057)
* [Veracity: its independency from false data](https://towardsdatascience.com/big-data-analysis-spark-and-hadoop-a11ba591c057)

[More reading](https://towardsdatascience.com/big-data-analysis-spark-and-hadoop-a11ba591c057)

**Different stages of Data Analysis**

Diagram

Description automatically generated

Shared Folder Slides, Big Data, Slide number 18

Gathering the data: Finding a data source and selecting and evaluating the data to be extracted. Here ETL or Extract, transform and load might be used to collect data from various sources.

Preparing the data: Structuring the data into a useable form

Choosing a model: choosing an appropriate analysis technique

Analysing the data: testing the model against the data to see if it holds

Presenting the results: Explaining the outcome to stakeholders

Making decisions: new direction based on the outcome

Data mapping techniques

Cloud mapping and transformation tools can identify and link

fields in two different data sets. This can be done to link several

identity numbers for a centralised database. Data types and data

size have to be compatible between the different sets as a source

field and destination field from each database are decided for

mapping. After the mapping of the fields or columns testing with

sample data shows if a data migration can be automated. Updates

are needed to maintain the identity after new inputs to the

databases.

## Common concepts in Big Data

Relational databases contain cells in rows containing the records for each object as columns or fields are used for measures or values to each category. The data is processed in flat files without hierarchy in Flat File Databases, but Relational Databases store the relationship between the data sets and add value to the information.

Python is one of the most useful programming languages for managing Big Data using dictionaries, lists, tuples, and sets.

Data at rest defines data stored at a physical location while data in motion is referring to rapidly changing data requiring real-time processing. Onsite computing solutions use database servers, and cloud computing solutions are based on scalable networks of distributed data.

## [Diagram Description automatically generated](https://www.geeksforgeeks.org/difference-between-algorithm-and-flowchart/)

Access further reading by clicking on the image above.

## The difference between an algorithm and a flowchart

* An algorithm resolutes a problem while a flowchart shows the flow of data
* Algorithms can be complex to construct while flowcharts are made easy to understand
* Strings are used in algorithms as opposed to shapes and symboles in flowcharts
* On debugging flowcharts are harder than algorithms

## Diagram Description automatically generated

## CHAPTER 2

## IT Laws and Regulations

The expansion of use of technical devices, from personal computers to IoT sensors and the amount of data available has been regulated by a set of protective laws. These laws are concerning users’ data, users’ rights, definition of cybercrime and health and safety regarding IT.

## Phishing and Social engineering attacks

An effective way to bypass the security controls put in place to protect organisations is to ‘[trick](C://Users/User/OneDrive%20-%20Just%20IT/Documents/BOOTCAMP/WEEK%201-Data%20Structures/LESSONS/Defending%20Against%20Phishing%20&%20Social%20Engineering%20Attacks.pdf)’ data owners into trusting the attackers. This will result in leaking of crucial data which will identify an attacker falsely as for example the owner of an account. The users will reveal their protected data believing to aid legit professionals. The stronger this belief is the longer the attacker has before a security alarm is made.

## The CIA triad

Organisations use the [CIA Triad](https://www.techtarget.com/whatis/definition/Confidentiality-integrity-and-availability-CIA#:~:text=Confidentiality%2C%20integrity%20and%20availability%2C%20also,with%20the%20Central%20Intelligence%20Agency.) model to guide policies to implement information security in the work place and for their users. The CIA Triad is consisting of:

* Confidentiality of information
* Integrity of data
* Availability of resources

## Cloud Security Controls

There are four common [cloud security controls](https://www.pcidssguide.com/cloud-security-controls-what-you-need-to-know/) in place to protect the cloud.

1. Deterrent: warning off potential attackers
2. Preventive: protecting weaknesses and vulnerabilities
3. Detective: identifying breaching attempts
4. Corrective: getting the data protection back to acceptable levels

## 

## Related Legislation

**Computer Misuse Act 1990**

Prohibits breaches of security regarding

* Accessing someone else’s device without permissions
* Committing further crimes
* Making changes or destroying software and hardware

Amendments made to the Computer Misuse ACT by the Police and Justice ACT 2006 which bans designing, distributing or possession of hacking tools, DOS and DDOS.

**The Copyright(computer Programs)regualtions 1992**

Replaced Copyright, Designs and Patent ACT 1998

* Rights for computer products
* Licensed software is only used by permission from the owner
* Protective restrictions made on Open Sourced material

**Data Protection ACT 2018**

Regulating all commerce within the UK while a similar new law, GDPR, The General Data

Protection Regulation 2018 replaced it for companies in business with EU countries.

Data Protection ACT 2018 contains details about:

* Defining transparancy of data content
* Purpose and use of data
* Timeliness of organisational data
* Data protection
* User rights

**Consumer Rights ACT 2015**

Digital and other products must meet the following standards:

* Fully functionig
* Functionig as designed and described
* Coming with a description

**Copyright, designs and Patents ACT1998**

This law ensure the rights of the designer, author or creator of published texts, dramatical or musical prodductions or otherwise artistic works to decide how and to what purpose and gain it can be used. This also entails makers of digital-design productions.

* Copy rights
* Plagiarism
* Damemges

**The Health and Safety( Display Screen Equipment) Regulations 1992**

## Regulations protecting workers at the workplace ensure everyone’s health and safety.

* Risk assessments must be made to recognize the risks.
* Recess allowance
* Yearly allowances for costs relating to eye examinations.
* Issues relating to poster and body alignment during work
* Health and safety training

## CHAPTER 3

## Excel

This section will present the making of a searchable drop-down list which includes the following functionalities:

a. It is possible to enter a letter to search for a name from a list of names containing the letter.

b. It has a dynamic name range which means changes to the name list will be updated to the drop-down list.

[WATCH THE VIDEO](https://www.youtube.com/watch?v=4wr-ns5vvpM)

TASK 1 A Searchable List

Start by selecting a column with the data you want in the drop-down list. Next, insert an empty column and call it “Helper”. Let the header to the data column be “Names”. Convert the columns into a table.

A picture containing application

Description automatically generated

For A Searchable list

1. Choose any two empty cell in the worksheet. Let it be cell J4.
   1. Format J3 and J4..



1. In the first cell of the “Helper”, A2, enter:

=SEARCH($A$2,Table2[@[Name]])

1. Add

ISNUMBER(……….)

* 1. The formula will be:

=ISNUMBER(SEARCH($A$2,Table2[@Name]]))

Now, when you enter a value into J4, the helper will show a TRUE or FALSE value to show if the letter is in any of the names.

1. Add IF(……….,1,0)
   1. The formula will be:

=IF(ISNUMBER(SEARCH($A$2,Table2[@[Name]],1,0)))

This will turn the TRUE or FALSE value into 1s and 0s.

1. Replace 1 by MAX($A$1:A1)+1, 0
   1. The formula will be:

=IF(ISNUMBER(SEARCH($A$2,Table2[@[Name]], MAX($A$1:A1)+1, 0) Now, the helper columns shows a serial number belonging to the position of the letter in each of the matching names in the name column.

A picture containing table

Description automatically generated



To complete the searchable list we need a row function to extract the matching names.

1. Let’s choose an empty column.
   1. In the fist cell, L2, enter

= ROWS($L$2:L2)

In ($L$2:L2), the first part, $L$2 is fixed but L2 in the second part is moving as the value entered is changed. We drag and drop the cell to map this column to the helper column. It will now show a serial number for each cell.

1. Add VLOOKUP(….., Table2, 2, FALSE)

To extract the names into the new column we use, VLOOKUP, enter the table name or array and then for the column number 2, and FALSE for exact matches.

2.1 VLOOKUP(ROWS($L$2:L2),Table2, 2, FALSE)

1. Add IFERROR(……,””)

3.1 IFERROR(VLOOKUP(ROWS($L$2:L2),Table2, 2, FALSE),””)



We now have our Searchable list and can move on to Dynamic ranges.

TASK 2 Setting the Range

1. We start by selecting a column for counting how many values there are.
   1. Enter into the first cell of the M-column

=COUNTIF($L$2:$L$38, “?\*”)

L38 is the last cell in the column showing the matched values, and ?\* is to mark any values in the L-column. This count will be for the total number of matches.

1. Add OFFSET ($L$2,,, COUNTIF($L$2:$L$38, “?\*))
2. Now create a name range by

Formula Name Manager New Name

DR\_NAME in the range click and paste

OFFSET ($L$2,,,COUNTIF($L$2:$L$38, “?\*))

Last part is to create a Front End to our Back End construction.



1. Format a cell in a new worksheet and call it Name.
2. Create a drop-down list by

Data Data Validation List and Press F3 on Source,

click on DR\_NAME click on Error Alert and untick “ Show Error

alert after invalid data is entered”

One last task is to link the drop-down list to the search cell in the previous worksheet.

Do the following:

Clear the search cell, enter = and go to the front end and click on the first cell under “Name”.

Table

Description automatically generated

Now, in the Front End worksheet enter a letter and see a dynamic range of names on a drop-down list to choose from.

Graphical user interface

Description automatically generated



As the letter J is entered into the search box, there is a number 1 indicating a match on the first letter of the text inside the cell and the name James Sullivan appears in the list column(L-column).



[See the List Here](https://justit831.sharepoint.com/:x:/r/sites/SkillsBootCamp-SoftwareDeveloper21-22/_layouts/15/Doc.aspx?sourcedoc=%7B816A9972-062E-462E-B6D1-318E15594875%7D&file=Supplier%20sheet.xlsx&action=default&mobileredirect=true)

**Charts are created and formatted in the Visualizations panel.**

Funnel chart: Click anywhere on the canvas, click on the columns to add to the chart in the Fields on the right-side panel, click online chart in the visualizations panel next to Fields and stylize by selecting format your visual and then general. Colour the bars by selecting them and click on visual and colors.

**Zoom Slider will add precision to filtering.**

Add a zoom slider to the chart, here to the line chart in the visualizations, Visuals, and turn the Zoom slider on.

Chart, funnel chart

Description automatically generated

Select the canvas, choose the columns, select the chart type, and stylize in the visuals panel. The above Screen shots show a Funnel chart to the left and a Line chart to the right.

Graphical user interface, application

Description automatically generated

[See the Dashboard here](https://teams.microsoft.com/_#/pbix/viewer/teamsSdk/https:~2F~2Fjustit831.sharepoint.com~2Fsites~2Fmsteams_5c1c0f~2FShared%20Documents~2FGeneral~2FAssignment%20Dashboard-Power%20BI.pbix?threadId=19:eKO3bewQwCt0fByFpPZamPt00wI-frfj711jF4dB6vA1@thread.tacv2&subEntityId=%257B%2522viewParams%2522%253A%2522id%253D%25252Fsites%25252Fmsteams%25255F5c1c0f%25252FShared%252520Documents%25252FGeneral%2526viewid%253Df4397a02%25252D0226%25252D4277%25252Daeed%25252Dbba078773ce8%2522%257D&baseUrl=https:~2F~2Fjustit831.sharepoint.com~2Fsites~2Fmsteams_5c1c0f&fileId=31b46e00-fd94-4a0c-9e18-2e0bc2f5191a&ctx=openFilePreview&viewerAction=view)

## CHAPTER 6

Dashboard in Tableau

Start by opening a worksheet for each component or chart in the dashboard.

Worksheet 1 Map

Drop the Date and Longitude column in the columns and put the latitude in rows. These columns are found in the Tables panel on the left hand side of the main panel. Drop a Sum of Unit Price in Colors and one in the Size containers. The State and City columns will be in the Details container.

Map

Description automatically generated

Worksheet 2 Horizontal Bar chart

Create a bar chart by dropping the category column in the rows and Sum Sales amount in the columns. Right click on the Sum sales column and hover on Measure and then to Average. This will change the aggregation level to Averages. Then Drop another Sum sales into the Details container and by right clicking find attributes. Make sure there is a copy in filters too.

Bar chart

Description automatically generated with low confidence

.

Worksheet 3 Packed Bubbles

Make sure the category column is in Labels and colors and the sum of sales column is in the size container

Chart, bubble chart

Description automatically generated

Lastly, from the Dashboard’s tab click on New Dashboard. The worksheets are on the left side and can be clicked into the sheet. Change the size of the dashboard by a jousting the Size box on the left-hand side. Change and format the name of the chart by double clicking on it in the dashboard.

Chart

Description automatically generated

[See the Dashboard here](https://public.tableau.com/app/profile/parisa.azimivahdat/viz/Assignment-Dashboard_16764068773900/Dashboard4?publish=yes)

## SQL Problem-Solving

6.

SELECT ProductID, ProductName, Price as 'Christmas Discount Price'

FROM Products

WHERE ProductName NOT IN('Ikura','Chang', 'Pavlova','Tofu')

ORDER BY Price ASC;

8.

SELECT OrderID, Quantity,

CASE WHEN Quantity <= 30 THEN 'Low Value Customers'

WHEN Quantity <= 50 THEN 'Value Customers'

WHEN Quantity <= 70 THEN 'Large Customers'

ELSE 'Premium Customers'

END AS QuantityText

FROM OrderDetails;

9.

SELECT Quantity \* Price as 'Sales'

FROM OrderDetails as 'odd'

JOIN Products as 'Prod'

on prod.ProductID=odd.ProductID

Where Sales >= 300;

10.

IN

11.

Case Statement

12.

Avoid Duplicate

13.

Order By

14.

LIKE