Data Superher

How to be a Data Analyst Booklet Guide

From zero to hero level with your Data Analyst Saver !

Written By : Reem Al-Ashhab
Revised : Fahad Masood Reda

Version 1

About the Author

Reem Alashhab is a Data Analyst – self learned, and she is at the same time a content creator in the field of Data Science. Her background is in Management Information Systems. She is very ambitious, passionate and enthusiastic Data Analyst known as a Data Analyst Superhero who uses the power of Python and R to get valuable insights from any data in her hands!

Table of Content

Introduction

I want to be a Data Analyst as soon as possible as everyone is quickly involved in this field posting their online certificates on LinkedIn !
A beautiful wish , but some really great wishes needs great time to become real ones !

I have heard the above wish so many times that I can't count ! Therefore , as my role is a Data Super Hero , I will take your hand out of all that chaos ! I will help you to become your own DATA SUPER HERO !

In this booklet ,I will briefly talk about the real characteristics of a data analyst and discuss the technical and soft skills of a data analyst. In addition, I will take you through the data analysis process and the core curriculum combined with courses and books to sharpen your analysis skills, and help you be to be your own data analyst super hero!

> *In the eyes of a Data Analyst, the world is an open book filled with data that if analyzed it will tell the story of life.*
>
> — Reem Al-Ashhab

Data Analyst , Data Scientist , Data Engineer

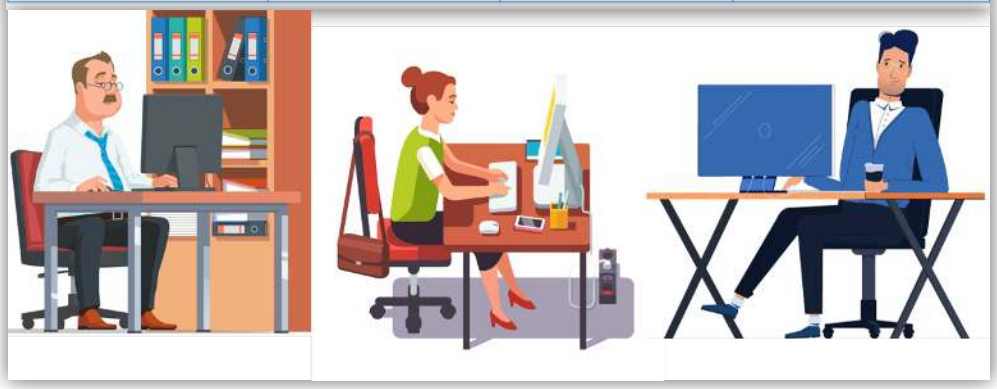
Who ARE THEY ?

A Data Analyst : is the one who is able to extract useful insights from a given data set. The data analyst can help companies to analyze market trends, find customers' needs, and asses their overall performance. A data analyst can summarize data and help top authorities to take decisions. The most used tools for a data analyst are Python, SQL and Excel.

A Data Scientist : is the one who uses his or her knowledge in statistics, programming, and machine learning to make predictions of the future. A data scientist is able to use machine learning algorithms to understand data patterns and create questions to be answered after data transforming and cleaning. The common tools for a data scientist are R, Python, SAS, SQL, and No-SQL.

Data Engineer : is the one who is able to gather or prepare data for a data analyst or a data scientist for further exploration and analysis. A data engineer is able also to create and maintain platforms for data processing . In other words, he or she can provide the best work environment for data various operations. The common tools for a data engineer are Java, SQL, No-SQL, Hadoop, and Apache Spark.

Data Engineer	Data Analyst	Data Scientist	Machine Learning Scientist
Collecting data from different sources and save it properly	Visualizing data and explore it using statistical methods	Extract insightful information from data	Make predictions and build models
SQL + No-SQL Java + Scala+ Python + Hadoop + Spark	SQL + BI Tools + Spreadsheets+ Excel Sheets	Python + R + SQL + No-SQL + SAS	Python + R



How data analysis is done ?

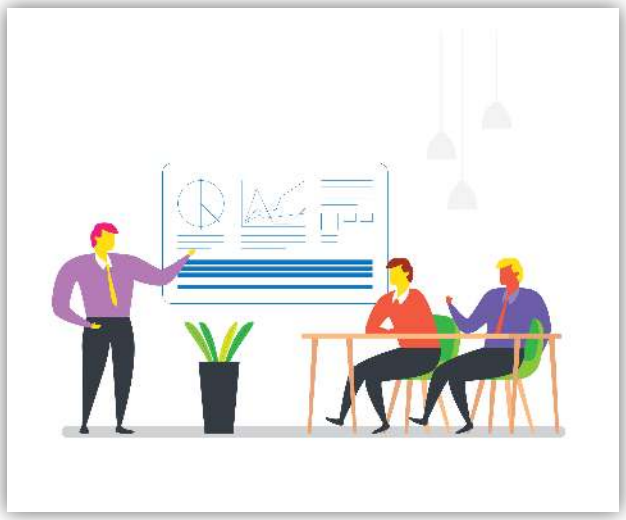
If you ever wondered what a Data Analyst do , you have to know the following :
It takes time to gather, asses and clean data, but the best part is when you extract the gold (Information) out of the deep ocean of data !

1. Data Gathering : Gathering data is the first step in the data wrangling process; it involves obtaining data (downloading a file from the internet, scraping a web page, querying an API, etc.), and then importing that data into a programming environment (e.g. Jupyter Notebook, R studio, Pycharm, and Spark).

2. Data Assessing : After gathering data required , the next step will be assessing this data. Data is inspected for two things: data quality issues (i.e. content issues) and lack of tidiness (i.e structural issues) before moving into the cleaning phase.

3. Data Cleaning : Cleaning data is the third step in data wrangling, where the quality and tidiness issues that were identified previously are solved by the help of a programming language like Python or R.

4. Wrangling and Analyzing Data : In this phase, the data that was previously gathered, assessed and cleaned will be visualized using the graphs and figures by the help of a programming language library (e.g. Python that includes libraries like matplotlib and seaborn).



What is the secret recipe of a data analyst ?

In this part, I want you to feel how it's like to be a data analyst .And understand what kind of a person is a data analyst? Is he or she a mathematician? A statistical scientist? or A programmer? I know that you want the secret recipe , so here it is !

· **Creativity:** When I first started to dig deep in the data science field, I was so worried of the variety of the things that I have to learn. There are many programming languages, many statistical methods, and many math equations. I asked myself how I'm supposed to learn and remember all of this?! And while I was working on different projects, I used to think of how to solve the challenge here. I started to google and look for answers everywhere. Then, I realized that there are so many solutions and so many tools to help me solve them. I reached the conclusion that You as a Data Analyst have to be flexible enough to do experiments and try different solutions to find answers. As time passes, you will have a lot of experiences that build your creativity in some way or another. So, remember a Data Analyst is a creative person who try and creates different ways to solve a problem!

· **Curiosity:** Webster said "At the end of the day, what you really want is someone who's curious, who enjoys drilling into data and spending a lot of time with it, hours on end, and coming up with a big insight that fuels the development of a strategy". While working on various data analysis projects, I found myself asking so many questions about the data in my hand, finding interesting answers , and then asking again new questions! I literally became a data detective! So, are you willing to be the same? Are you ready for a long journey of chasing the truth everywhere and every time? Does the possibility of failure scare you if you don't find the answers you are looking for?

· **Effective communication skills:** In one of my previous job interviews, I was presenting a data analysis project to several people in the company. On one slide, I was explaining a graph, and then one of the audience told me "I don't get the meaning of the graph! What are you trying to say?". At that moment, I realized that story telling is definitely a skill I should work on more! Zhang said "Storytelling is a very important skill—to [be able to] translate the numbers, the metrics, into very clear and straightforward, actionable and engaging stories". So, why should a Data Analyst care about communication? The simple direct answer is to make people trust your results! No one will trust your numbers and graphs if you don't know how to explain them well ! Therefore, remember communication, then communication and finally communication!

. Continues Learning: Through my learning journey, I faced many and many challenges I didn't know how to overcome. I put my effort and time to solve projects I have never done before. In all my work, I had to search and find articles, videos, and courses. Eventually, I found myself in a non-ending loop of continues learning. At first, I thought am I doing something wrong? Is this normal for any data analyst? And even in another days, I found myself forgetting how I solved a previous challenge! I had to go back through my previous work and study it again! Then I reached the conclusion that it's totally normal. A Data Analyst must be always hungry for more and more as data is growing more and more at every second!

· Statistical and technical expertise: Should I know statistics? What should I learn first Python or R? All these questions and others are very important to be answered before you dig deep in the data analysis and become an expert. Many people in this process will feel lost and eventually walk away because they can't clear this chaos. My answers for all the above questions are as the following: Yes, you have to learn statistics continuously! It's normal to forget what you learn, but do your best to build a strong statistical background. For the question, which programming languages should I start to learn? You should specify your objectives. Do you want to start a new career? If yes, then study the market very well. If you know R, you may not use it because the IT Department team in the company prefers Python and vice versa! (Note: this is just a self-made example, and not a real fact!). On the other hand, if you want to work as a freelancer, you will need also to study the market and find the gap that you can fulfill. There's no certain answer; it actually depends on what objectives you want to work on! Final note for you to remember: There is no perfect programming language! All programming languages serve different purposes, so please be very clear of what you are trying to achieve !

Study Plan for Data Analysis

At the beginning of your learning journey, you will find many fake ads that tell you that you can be a data analyst in one month, I'm telling you from now their claim is absolutely the opposite of the truth! I have spent nearly 6 months in data analysis, and I still feel that I don't know anything! Let me tell you why you will never feel that you know everything because simply the data science road doesn't have an end ! You have to develop your skills over each day of your precious life!

Therefore, let's go immediately over the study plan for data analysis! The provided plan is taken from Udacity Nanodegree for 'Data Analyst' and improved by me. This study plan is the best to follow from my experience as a beginner in data science field. However, the course itself is a little bet costly for some people. For this reason, I will provide also free courses as an option for people who can't afforded it, but let's us first evaluate the Nanodegree program.

Data Analysis Nanodegree Program

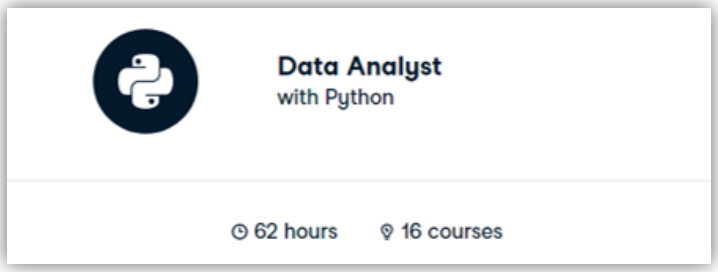
Anyone asking me " How did you start as a data analyst ?" I tell them I studied at Udacity. To be honest, I didn't feel lost while studying the data analyst nanodegree because the path of study and everything was very clear for me ! I tried other platforms such as Datacamp. Datacamp has other advantages that make it a really good platform for anyone interested data science. However, the best of the best stills the ' Data Analyst Nanodegree program' at Udacity ! Additional note for non-programmers, don't jump into this course before taking some programming courses first !

ESTIMATED TIME	PREREQUISITES
4 Months At 10 hrs/week	Fundamentals of programming

FEATURES	COST
1. Short videos. 2. External resources. 3. Short quizzes. 4. Real - world projects from industry experts. 5. Technical mentor support 6. Personal career coach and career services 7. Interview preparations 8. Resume services 9. Github review 10. LinkedIn profile review 11. Full path of study	Pay as you go : After your 7 days trial, pay 400$ /month OR Pay 1,439 $ for 4 months.
COURSE LINK	**COURSE SYLABULS**
https://cutt.ly/qgvIlZO	https://cutt.ly/igwAgBK

Datacamp

Datacamp is a really good alternative for Udacity. Datacamp has more diverse courses than Udacity. It also offers more than 80 projects related to data science to work on.



ESTIMATED TIME	PREREQUISITES
8 weeks At 8 hrs/week	Fundamentals of programming
FEATURES	**COST**
1. Video Lessons combined with real-time practices. 2. More than 80 real Projects to work on (I usually spend 1-2 hours only to complete a project). 3. Online practices and assessments to sharpen your skills. 4. Various articles related to Data Science. 5. Data Science Podcast.	Free: 0$ Standard : $156 per year. Premium : $ 399 per year.
COURSE LINK	**COURSE SYLABULS**
https://cutt.ly/ogvlYfb	https://learn.datacamp.com/career-tracks/data-analyst-with-python

365 Data Science

365 Data Science is definitely one of the best platforms to start as a Data Analyst besides Udacity . Their way of presenting includes motion graphic characters. If you subscribed to their platform , you will be led by them through all the path as both a data scientist and data analyst.

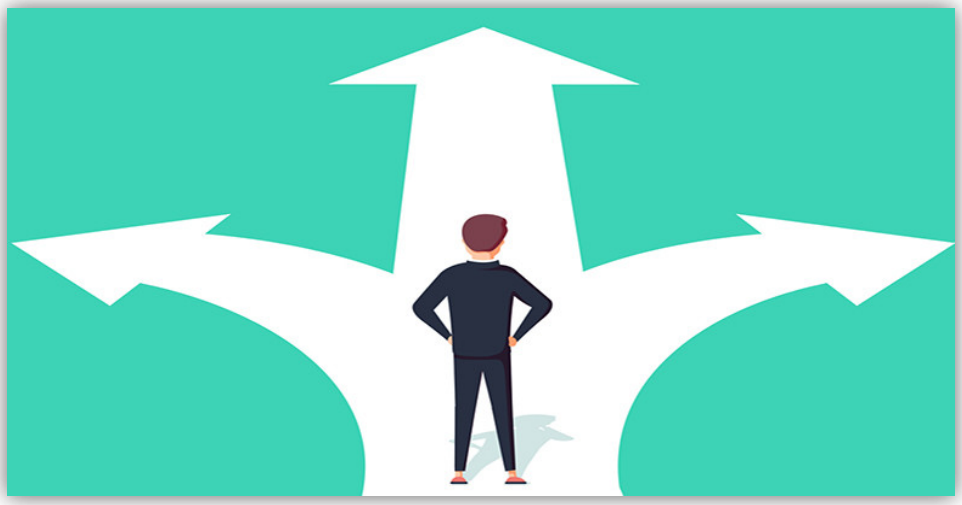
ESTIMATED TIME	PREREQUISITES
6-12 months At 10 hrs/week	Fundamentals of programming and Mathematics
FEATURES	**COST**
1. Learn from a structured curriculum. 2. Bring theory into practice. 3. Earn verifiable certificates.	After your 7 days trial, pay 36 $ /month or 348 $ per year

COURSE LINK	COURSE SYLABULS
https://cutt.ly/bgvlOj1	1. Intro to Data and Data Science 2. Intro to Microsoft Excel 3. Statistics 4. Probability 5. Introduction to Python 6. Mathematics 7. Intro to Programming 8. Tableau 9. Power BI 10. Advanced Statistical Methods in Python 11. Deep Learning with TensorFlow 12. Time Series Analysis with Python. 13. SQL 14. Advanced Microsoft Excel 15.SQL+Tableau 16. Git and Github 17. Machine Learning with Python 18. Deep Learning with TensorFlow 2.0 19.Customer Analytics in Python 20.Web Scrapping with Python 21. Python for Finance 22. Data Preprocessing with NumPy 23.Product Management for AI & Data Science 24.Starting a Career in Data Science: Project Portfolio, Resume, and Interview Process 25. Credit Risk Modeling 26. Introduction to Business Analytics

Data Analysis Plan of Study

In this section, I will present the map of study that's implemented by Udacity for the data analysis nanodegree program accombined with suggested free and paid courses.

1. Introduction to Programming.
2. Introduction to Python.
3. Introduction to SQL.
4. Database Design (Optional).
5. Practical Statistics.
6. Introduction to R (Optional).
7. Data Wrangling.
8. Data Story Telling.
9. Data Analysis in Excel. (Industry requirement).



Introduction to Programming

This section is purely for beginners in the programming field. Before diving in the data analysis field, you must learn how to use some tools first ! One of the best courses to learn fundamentals of **programming is CS50** . The course is presented by Oxford university ! So imagine how the course will be ! This course is the best course to start from zero to hero ! Check the course reviews online if you wish !

ESTIMATED TIME	PREREQUISITES
12 weeks Effort : 6–18 hours per week	None

Course Overview	COST
1. You will learn how to think algorithmically and solve programming problems related to Computer Science. 2. You will learn significant concepts such as abstraction, data structures, algorithms, encapsulation, security, resource management , software engineering and web development. 3. You will learn in a number of programming languages such as C, Python, SQL, and JavaScript plus CSS and HTML.	FREE Add a Verified Certificate for $90 USD

COURSE LINK	FEATURES
https://cutt.ly/kgvlVZu	1. Well explanation for programming concepts. 2. Projects to work on after each section or chapter. 3. Valid certificate from Harvard University.

Introduction to Python

Now, you will tell me why to learn Python and not R ? To answer your question , I recommend Python because it's very easy and highly requested in the job market while R Language is more into research and academics . However, in this booklet, you will find both options to choose your own path ! I recommend the Python course to be taken at Datacamp platform.

ESTIMATED TIME	PREREQUISITES
4 weeks Effort : 4 hours per week	Programming fundamentals

WHAT YOU WILL LEARN	COST
You'll learn how to import, clean, manipulate, and visualize data—all integral skills for any aspiring data professional or researcher. Through interactive exercises, you'll get hands-on with some of the most popular Python libraries, including pandas, NumPy, Matplotlib, and many more.	Free: 0$ Standard : $156 per year. Premium : $ 399 per year. For More details check : https://cutt.ly/LgeMu2y
COURSE LINK https://cutt.ly/ogvlYfb Note : your subscription will gain you access to more career paths such as python for data scientist . **COURSES TO BE TAKEN** 1. Python Fundamentals (select this option from Skills Tracks section).	**FEATURES** 1. Video Lessons combined with real-time practices. 2. Real Projects to work on (I usually spend 1-2 hours only to complete a project). 3. Online practices and assessments to sharpen your skills. 4. Various articles related to Data Science. 5. Data Science Podcast.

Introduction to SQL+ Database Design

You may now ask yourself : why to learn SQL and Database design ? isn't the job of a Data Engineer ? I will simply tell you that sometimes a Data Analyst might need to get data by himself or herself to be ready for any circumstances . Therefore, just learn them both at Datacamp under the Data Analyst with SQL Server

ESTIMATED TIME	PREREQUISITES
6-8 weeks Effort : 8 hours per week	Programming fundamentals

WHAT YOU WILL LEARN	COST
You'll learn how to import, clean, manipulate, and visualize data—all integral skills for any aspiring data professional or researcher. Through interactive exercises, you'll get hands-on with some of the most popular Python libraries, including pandas, NumPy, Matplotlib, and many more.	Free: 0$ Standard : $156 per year. Premium : $ 399 per year. For More details check : https://cutt.ly/LgeMu2y

COURSE LINK	FEATURES
https://cutt.ly/ogvlYfb Note : your subscription will gain you access to more career paths such as python for data scientist .	1. Video Lessons combined with real-time practices. 2. Real Projects to work on (I usually spend 1-2 hours only to complete a project). 3. Online practices and assessments to sharpen your skills. 4. Various articles related to Data Science. 5. Data Science Podcast.

Introduction to SQL

"' The Complete SQL Bootcamp 2020: Go from Zero to Hero "' course on Udemy is one of the top rated courses. The course is very good for beginners as you will find all concepts well explained and demonstrated.

ESTIMATED TIME	PREREQUISITES
1-2 weeks Effort : 8 hours per week	No prior technical experience is required
WHAT YOU WILL LEARN	**COST**
1. How to write complex queries to a database using PostgreSQL, which is one of the most in demand skills - PostgreSQL. Note : These skills are also can be used with other major SQL database : such as MySQL, Microsoft SQL Server, Amazon Redshift, Oracle, and much more.	$189.99
COURSE LINK	**FEATURES**
https://cutt.ly/Jgvl1Gt	1. 9 hours on-demand video 2. 15 articles 3. 10 downloadable resources 4. Full lifetime access 5. Access on mobile and TV 6. Certificate of completion

Database Design

If you are not interested in investing money in the course . "Database Systems Concepts & Design" course is the best option to learn all about database design , offered by Udacity .

ESTIMATED TIME	PREREQUISITES
8 weeks Effort : 8 hours per week	No prior technical experience is required
WHAT YOU WILL LEARN	**COST**
This course presents database fundamentals and database design concepts.	Free
COURSE LINK	**FEATURES**
https://cutt.ly/mgvl9U3	1. Rich Learning Content 2. Taught by Industry Pros 3. Interactive Quizzes 4.Self-Paced Learning

Practical Statistics

Learning Statistics is very important for any data analyst as data science is built on statistics and mathematics.

ESTIMATED TIME	PREREQUISITES
8 weeks Effort : 8 hours per week	No experience is required
WHAT YOU WILL LEARN	**COST**
The course will cover topics that will help you to understand statistics applied to data such as visualization, probability, regression..etc.	Free
COURSE LINK	**FEATURES**
https://cutt.ly/Ygvl5D1	Rich Learning Content Taught by Industry Professionals Student Support Community Interactive Quizzes Self-Paced Learning

Introduction to R

Although it's preferred to learn Python to meet companies needs, R is still an important language to learn to add more valuable skills to you resume .

ESTIMATED TIME	PREREQUISITES
10 weeks Effort : 8 hours per week	Programming Fundamentals

WHAT YOU WILL LEARN	COST
In this course, you will learn valuable skills required for data analysis using R such as data cleaning, manpulation, and visualization . In addition, you will learn how to use the following libraries : ggplot2, tidyverse, dplyr, and readr..	Free: 0$ Standard : $156 per year. Premium : $ 399 per year. For More details check : https://cutt.ly/LgeMu2y

COURSE LINK	FEATURES
For More details check : https://cutt.ly/7gv0f8b **COURSES TO BE TAKEN** Under skill section study : R Programming Importing & Cleaning Data with R Data Manipulation with R	1. Video Lessons combined with real-time practices. 2. Real Projects to work on (I usually spend 1-2 hours only to complete a project). 3. Online practices and assessments to sharpen your skills. 4. Various articles related to Data Science. 5. Data Science Podcast.

Data Wrangling

Data wrangling means going through the phases of data analysis , which are data gathering, data assessing data cleaning, and data exploratory.

ESTIMATED TIME	PREREQUISITES
10 weeks Effort : 8 hours per week	Programming Fundamentals of Python
WHAT YOU WILL LEARN	**COST**
You'll learn how to import, clean, manipulate, and visualize data.	Free: 0$ Standard : $156 per year. Premium : $ 399 per year. For More details check : https://cutt.ly/LgeMu2y
COURSE LINK For More details check : https://cutt.ly/7gvOf8b **COURSES TO BE TAKEN** Introduction to Data Visualization with Matplotlib Introduction to Data Visualization with Seaborn Introduction to Importing Data in Pyhton Intermediate to Importing Data in Python Cleaning Data in Python Exploratory Data Analysis in Python	**FEATURES** 1. Video Lessons combined with real-time practices. 2. Real Projects to work on (I usually spend 1-2 hours only to complete a project). 3. Online practices and assessments to sharpen your skills. 4. Various articles related to Data Science. 5. Data Science Podcast.

Data Story Telling : Power BI

Learning how to make great and beautiful diagrams through Power BI is an essential skill for any data analyst. Enjoy this top rated course ! Microsoft Power BI - A Complete Introduction [2020 EDITION].

ESTIMATED TIME	PREREQUISITES
3 weeks Effort : 8 hours per week	Basic understanding of data analysis is a plus but not required
WHAT YOU WILL LEARN	**COST**
You'll learn how to best use Power BI for data visualization purposes .	175 $
COURSE LINK	**FEATURES**
For More details check : https://cutt.ly/egvOkES	1. 23.5 hours on-demand video 2. 26 articles 3. 127 downloadable resources 4. Full lifetime access Access on mobile and TV 5. Assignments 6. Certificate of completion

Data Story Telling : Power BI

ESTIMATED TIME	PREREQUISITES
1 week Effort : 4.5 hours per week	Basic understanding of data analysis

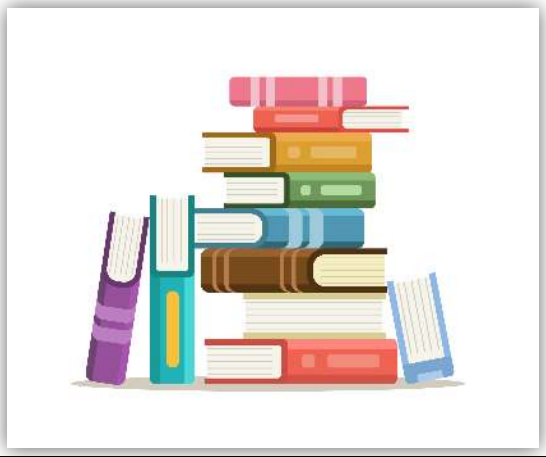
WHAT YOU WILL LEARN	COST
You'll learn how to best use Power BI for data visualization purposes . **1. Connect Microsoft Power BI to data sources** **2. Create Barcharts** **3. Create Treemaps** **4. Create Donut Charts** **5. Create Waterfall Diagrams** **6. Create Piecharts**	150 $

COURSE LINK	FEATURES
For More details check : https://cutt.ly/VgvOxul	1. 4.5 hours on-demand video 2. 4articles 3 Full lifetime access Access on mobile and TV 4. Certificate of completion

Data Analysis in EXCEL

ESTIMATED TIME	PREREQUISITES
1 week Effort : 3 hours per week	Basic understanding of data analysis
WHAT YOU WILL LEARN	**COST**
You'll learn how to use excel for data analysis purposes	FREE Add a Verified Certificate for $90 USD
COURSE LINK	**FEATURES**
For More details check : https://cutt.ly/6gvOnqH	1. Shareable Certificate 2. Flexible deadlines

Suggested Books in Data Analysis



The Art of Data Science https://cutt.ly/ngvOQPX	Data Science from Scratch https://cutt.ly/OgvOWNt
Head First Data Analysis: A learner's guide to big numbers, statistics, and good decisions https://cutt.ly/ngvORnC	HBR Guide to Data Analytics Basics for Managers (HBR Guide Series) https://cutt.ly/lgvOlqg

Conclusion

This is the first version of " How to be a Data Analyst ?". There will be monthly updates on this booklet. My ambition is to turn this simple booklet to a whole book in the near future.

References

https://cutt.ly/hghUGLB
https://cutt.ly/TghUCFS

Thank You

I would like to thank you all for reading my first booklet in Data Analysis. This is my first attempt to draw a clear path for anyone interested to enter the data science field and be a data analyst. Thank you again , and wait for the second version of this booklet !

Contact me on :

LinkedIn Profile
https://www.linkedin.com/in/reemalashhab/

Email Address
alashhab23@gmail.com