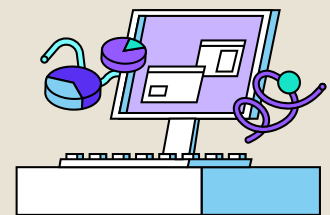




# SYLLABUS DATA ANALYST



**DISCOVER THE DIFFERENT JOBS IN DATA**

# OUR STORY

DataScientest is the leading training organization in Europe for Data Science. We offer business-oriented training courses for upskilling and reskilling to professionals and individuals alike.

## TRAINING DESIGNED WITH AND FOR COMPANIES

Data has become the main resource to exploit for companies. It allows them to **guide their decision-making** and to predict the behaviour of their environment. Therefore, staying **competitive** means exploiting data to its fullest potential. Thus, given that data is everywhere, it is essential to **increase workforce capabilities** on these new skills, tools and technologies to ensure stable development.

This is why DataScientest has been training the Data teams of more than **70 Fortune 500 groups** as well as many other SMEs and Startups for over **6 years**. Since 2021, more than **5500 professionals** have been trained with us.

# 74%

of the heads of Data in the largest groups wish to strengthen their teams by recruiting Data Analysts\*

\*Survey conducted among our forty partner companies.

## A CURRICULUM ADAPTED TO INDIVIDUALS

Having been designed for our **partner companies**, our training courses are professionally oriented, as close as possible to the **needs of the market**. Thus, in order to face the shortage of Data profiles and bridge this gap, we have decided to open our training courses to **individuals**. Under this impulse, DataScientest is constantly developing thanks to **new training courses** (Data Manager) or **new specialized courses** (Deep Learning).



*"DataScientest is a rich, personalized learning experience that responds to the real problems of the company."*

Samir Ait Idir  
Chief Data Officer & Head of Data Intelligence @Orange Bank

# DATASCIENTEST KEY FIGURES

**+70** FORTUNE 500 COMPANIES  
AS CLIENTS

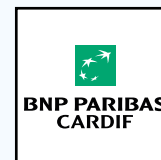
**98%** COMPLETION  
RATE

**94%** SATISFACTION  
RATE

**+5500** ALUMNI

**2000h** OF EXCLUSIVE  
CONTENT

## THEY TRUST US





# OUR PARTNERS

DataScientest has developed partnerships with academic institutions on the one hand and with software publishers on the other. The content of our training courses on Business Intelligence or Cloud tools is henceforth certified by the companies that own them. Here are our different partnerships:



**AMAZON WEB SERVICES**  
Editor partner

Today, DataScientest enjoys the exclusive status of **Amazon Training Partner**. We are therefore authorized by Amazon to train teams on the products and services of the American company. Thanks to this partnership, we have established **several training courses** that prepare for **official AWS certifications**. The registration fee for the official exam is included in the price of the course.

In some of our courses, you will be prepared to take the fundamental **AWS certification: Cloud Practitioner**.



**MICROSOFT - POWER BI**  
Editor partner

DataScientest is a **Microsoft Learning Partner**, which means that we can train you for **official Microsoft certifications**. These certifications attest to a level of expertise in Azure, the set of cloud computing products and services, and in **Power BI**, Microsoft's business intelligence tool. The registration fee for the official exam is included in the price of the course.

In each training courses, DataScientest prepares you to pass official Microsoft certifications. Additionally, with the **Data Management** course, DataScientest allows you to prepare for the **PL-900** certification, **Microsoft Certified Power Platform Fundamentals**.



**HEC DATA MINDS**  
Academic partner

Since October 2020, DataScientest has been providing **HEC Paris** students with access to its platform to master coding and analytical problem-solving through data science.



**QUALIOPi**  
Quality criteria

This label recognises that the processes implemented by the training organization meet the quality requirements demanded by the State. **All DataScientest training courses are Qualiopi certified**.

# OUR TEAM

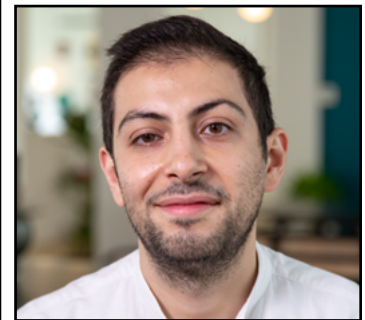
## OUR TEACHERS

DataScientest's teaching team is made up of internal professors. They are dedicated to teaching and research for our various courses and expert curricula.

All our professors are always there for you. They have created and continue to update all our courses. They will accompany you by video-conference and on our platform throughout your training.

With their excellent academic background and varied professional experience, they are the Data Science experts who will enable you to join the Data team in the sector that interests you most (Banking & Insurance, Industry, Finance, Medical and many others).

**Charles S.**  
Head of Academic  
*8 years of expérience*



A graduate of the École Polytechnique, Charles has been involved in the development of the courses from day one. He is specialized in programming, Machine Learning and Deep Learning.

**Raphael K.**  
Pedagogical Director  
*8 years of expérience*



Holder of the ISF master's degree specializing in Statistical Learning and Data Science from the University of Paris-Dauphine, Raphaël designed the Data Analyst and Data Management courses thanks to his knowledge in programming, dataviz and Machine Learning.

## OUR TEAM

**Lara E.**

Data Analyst course referent

*2 years of expérience*



Lara holds a master's degree in economic and financial statistical modelling from the University of Paris I Panthéon Sorbonne. A few years ago, she joined the DataScientest team as a Data Scientist. She quickly specialised in Machine Learning and is now in charge of the Data Analyst training.

**Thomas B.**

Deep Learning course referent

*6 years of expérience*



After attending a top tier engineering school, Thomas quickly entered DataScientest's ranks. His array of knowledge stretches from programming and Dataviz to NLP, Deep Learning and Computer vision.



**DANIEL**

During your training, you will progress on our platform. However, **you will never be alone**, Daniel is there for you. All our teachers and Program Managers take turns to **answer all your questions live via Slack** and on a dedicated support platform. With support every day of the week, from 9am to 5pm, you are guided throughout your course.

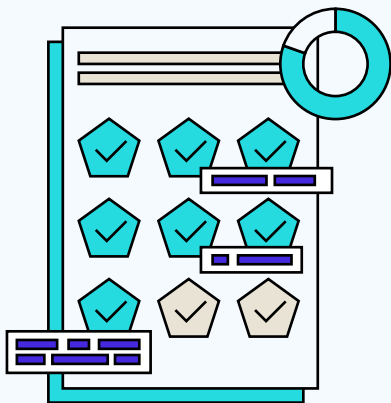
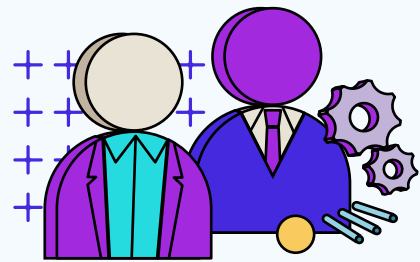


# OUR PEDAGOGY

## HYBRID FORMAT

DataScientest offers 100% distance learning in a hybrid format :

During approximately 20% of the duration, you are accompanied by your teacher and your cohort during **coaching sessions via videoconference**. The rest of the time, you work on our teaching platform and are **guided** via Slack by our **Data Scientists**. This format gives you the flexibility to organize yourself. We also monitor your progress and accompany you to ensure the **successful completion** of your training.



## MASTERCLASS

Each training period on the platform (called a Sprint) is accompanied by one or more **masterclasses by videoconference**. The objectives and teaching formats of each masterclass change according to the sprint: correction of concrete use cases, lectures on specific topics, competitions between cohorts, etc.

## EXAMS

At DataScientest, there are no automated MCQs. **Each paper is reviewed and corrected** by hand by our teachers. They go over all your difficulties with you.

### BACKBONE PROJECT

From the beginning of your training, you will carry out a concrete project with the aim to deploy it. It requires an investment of about 80 hours of work throughout the training.



You will work in pairs or in triads on a project that you will select from our regularly updated **project catalogue**. Our subjects are inspired by the work we do with companies.

This is a crucial step in your career path that will make you fully **operational**. You will be working on un-cleaned data sets, and will be expected to produce work of professional quality.

Reviews are done at regular intervals by your teacher to guide you, and **coach you**. This allows you to move efficiently from theory to practice and to ensure that you master the skills required on the different modules.

It is also a project that is **highly valued by companies**. It confirms your skills and knowledge acquired at the end of the **Data Analyst** course. You will then be able to justify your skills with a **successful Data Science project** during your interviews



*DataScientest is a rich, personalized course, perfectly adapted to today's corporate data science topics.*

Xavier Bocher  
Head of Credit Risk Internal Models & Operational Research @Groupe Crédit Agricole



## OUR PEDAGOGY

### PREREQUISITES

Our Data Analyst training is demanding and requires a good knowledge of mathematics. A 2-year degree with a business or scientific component is therefore required. It is essential that you discuss with our training team whether your project is relevant to our training.

### CONTINUOUS FORMAT

Only one format is available for the Data Analyst training :



## CONTINUOUS FORMAT

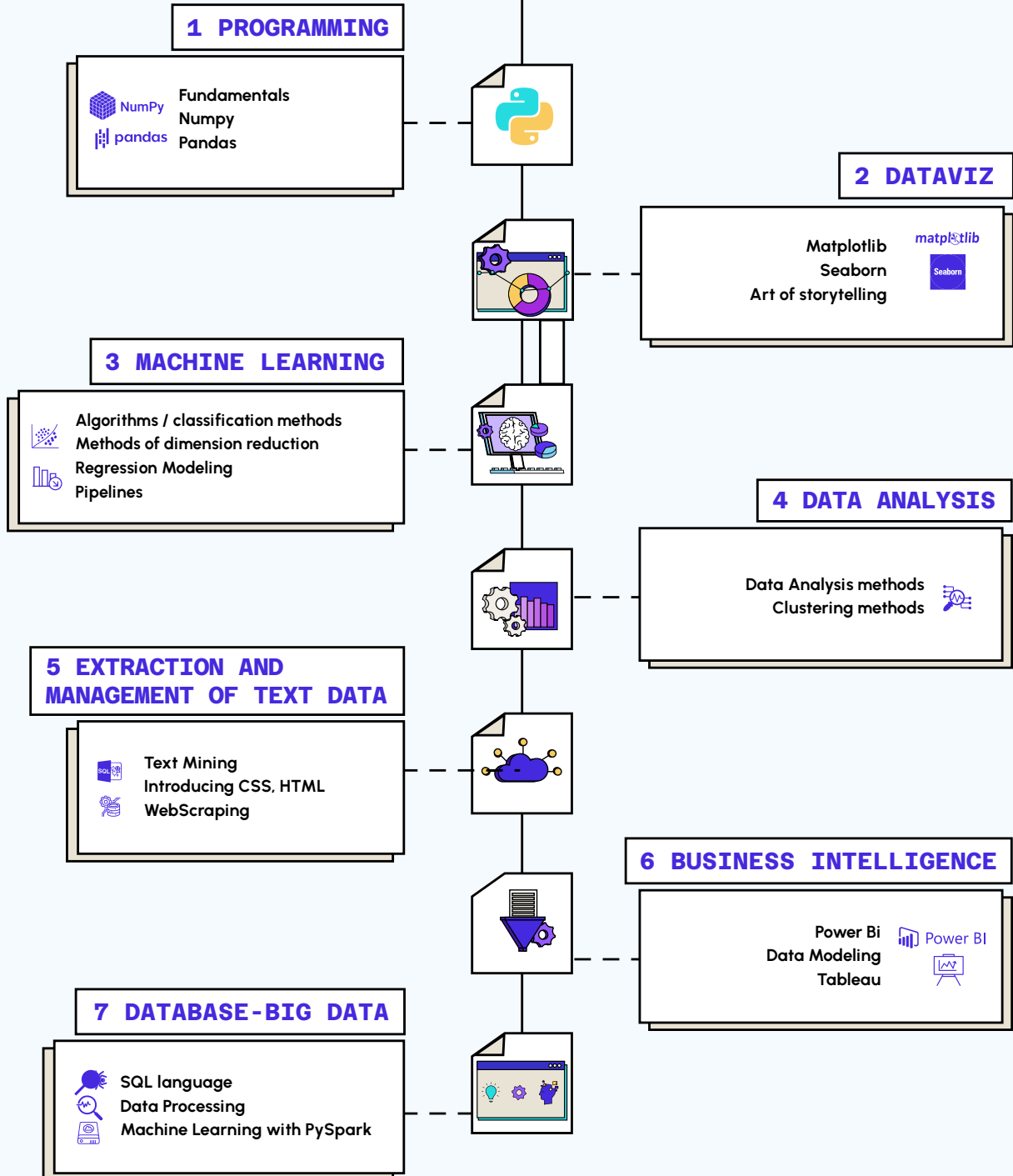
Flexible, it is the format adapted for the practice of a profession.

**DURATION** 6 months

**RYTHM** 10-12h / week

# CURRICULUM

## Data Analyst



## INTRODUCING PYTHON - duration 50h

### Fundamentals of Python

- Discovery of the different variables, lists and tuples
- Presentation of the various operators and control structures
- Introduction to the concept of loop and its different types
- Functions and documentation
- Instantiation of classes and use of modules

### Numpy

- Creation and manipulation of a Numpy Array
- Presentation of matrix operations
- Creation of a statistical indicator and applications

### Pandas

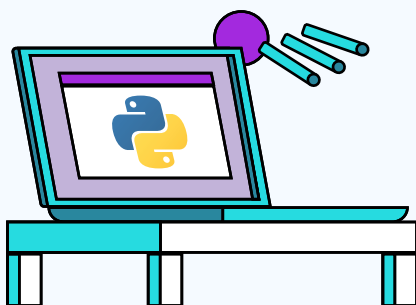
- Loading and first exploration of a dataset
- Data Cleaning
- Data Processing

### Data Quality (optional)

- Presentation of the essential principles of Data Quality
- Treatment of numerical and textual variables
- Data cleaning
- Managing missing values

### Skills acquired at the end

- Read and understand a Python code, the reference language in Data Science
- Handle and manage data tables
- Interrogate, manipulate, order and modify a dataset with Python



## DATA VIZUALISATION - duration 25h

### Matplotlib

- Presentation of different types of graphs:
  - Curves
  - Graphs
  - Scatter plot
  - Histograms
- Chart customization

### Seaborn

- Distribution analysis
- Statistical analysis
- Multivariate analysis

### Complément add-on (optional)

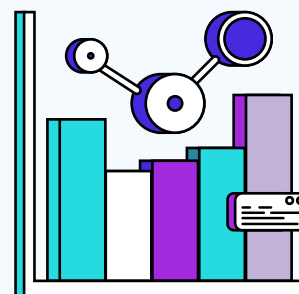
- Presentation of advanced graphics with Matplotlib
- Creating figures with classes and objects

### Bokeh (optional)

- Training in all types of graphics interactive and embeddable on a web page

### Skills acquired at the end

- Mastering and customizing a wide variety of fundamental graphic renderings for Data Visualization
- Be able to use Data Visualization for data analysis
- Production of simple statistical graphs at the crossroads of Data Visualization and data analysis
- Master the best practices of Data Visualization and Data Storytelling
- Create an interactive graph and integrate it to a web page



## MACHINE LEARNING - duration 50h

### Algorithms and methodology for classification with Scikit-Learn

- Presentation of classification algorithms (Logistic regression, kNN, Decision tree, Random forest, SVM...)
- Boosting and Bagging algorithms
- Model selection
- Classification of unbalanced data

### Dimension reduction methods (optional)

- Introduction to Principal Component Analysis
- Presentation of the T-SNE algorithm
- Introduction to discriminant linear analysis
- Discovery of clustering with the K-means algorithm

### Regression methods

- Simple and multiple linear regression
- Regularized linear regression (Lasso, Ridge and Elastic Net)

### Skills acquired at the end

- Master data analysis techniques and multivariate statistical analysis
- Implement a classical Machine Learning algorithm
- Preprocessing of data before modeling the data
- Evaluate a model using cross validation and various metrics
- Mastering boosting and bagging algorithms widely used in Machine Learning
- Identify unsupervised Machine Learning problems
- Use the main clustering algorithms using the scikit-learn library
- Know the performance metrics associated with clustering issues
- Master the simple linear regression model
- Apply different regression models

## DATA ANALYSIS - duration 20h

### Data Analysis

- Principal Component Analysis
- T-SNE
- Linear Discriminant Analysis (LDA)
- Clustering with the K-means algorithm

### Skills acquired at the end

- Mastery of the main methods of data analysis
- Mastery of the main algorithms of clustering with the Scikit-Learn library

### Clustering methods (optional)

- K-means algorithm
- Hierarchical Ascending Classification Algorithm
- Mean Shift Algorithm





## EXTRACTION AND MANAGEMENT OF TEXT DATA - duration 25h

### Text Mining

- Introduction to regular expressions
- Managing textual data
- Creation of wordclouds
- Sentiment analysis

### Webscraping

- Introduction to web language (HTML, CSS)
- Web content extraction with BeautifulSoup
- Application of scrapping on Google

### Skills acquired at the end

- Master the techniques of manipulation and preprocessing of text data
- Read and use regular expressions
- Use a Machine Learning model on textual data
- Systematically retrieve data from various various web pages



## BUSINESS INTELLIGENCE - duration 25h

### Business Intelligence

- Introduction to BI: concepts and usage

### Power BI

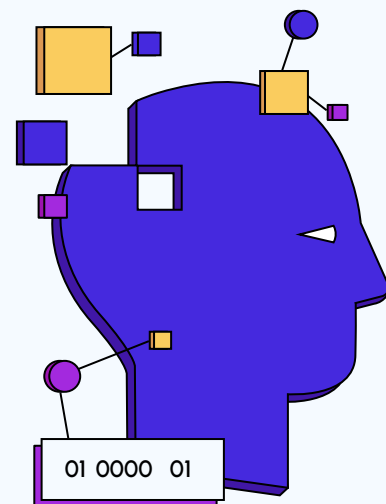
- Connection to data sources
- Data Visualization with Power Query
- Relationships and Data Modeling transformed Data
- Introduction to the DAX language
- Creation of an in-depth dashboard and case study

### Tableau

- Connection to data sources
- Data Formatting
- Data Visualization

### Skills acquired at the end

- Master the fundamentals of Power BI
- Use advanced tools (UML, Power Query, Dax...) in Data Analysis
- Visualize data at a professional level
- Learn and use the Dax language



## BIG DATA / DATABASE - duration 25h

### Introduction to Data Engineering and Big Data

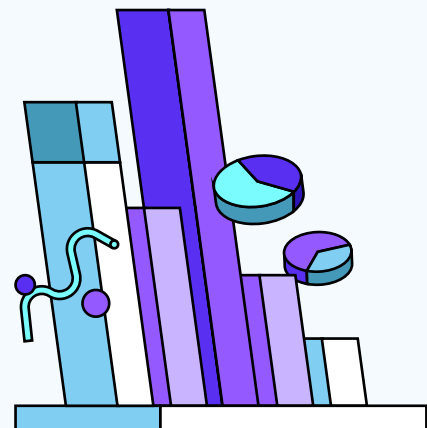
- Introduction to the 3V of Big Data (Volume, Variety, Velocity)
- Presentation of databases and Architectures
- Putting data into production

### Data Process and Machine Learning on large databases

- Introduction to PySpark
- Discovery of the different functions of PySpark
  - Data Processing
  - DataFrames
  - Regression with PySpark
  - The ML pipeline
  - Model Turning

### Skills acquired at the end

- Read and query relational databases
- Master the syntax of SQL queries
- Training in processing massive datasets using distributed computing
- Efficiently apply Machine Learning models to large databases



# ALUMNI TESTIMONIES



## **Amir ABBASZADEH**

Cofounder and CEO at Oilifina Consulting

Trained as an engineer, Amir started his career in strategy consulting at McKinsey and then in M&A at an investment bank. Following an entrepreneurial experience, he felt a real technical need in data processing but his 100% business profile remained a brake to launch himself into the adventure.

“DataScientest has completely erased this worry in my head and allowed me to realize my project. For me, what distinguishes DataScientest from others is above all the perfectly operational content and the live learning method! This always makes you want to continue and no one gets bored in the middle of the course. The bonus of this training is the incredible availability of the experts to help us. Today I have created my company thanks to this training and we are 8 people 8 of which 3 Data Scientists.

## **Pauline SCHIEVEN** Business Analyst - Revenue Operations at Jellysmack



Pauline holds a Master's degree in International Marketing and Business Development. She worked as a Territory Account Manager at UberEats.

Wishing to specialize in Data and to make it her future job, Pauline chose to trust DataScientest for her career project. Following our recruitment fair organized in November 2021, she found a job at Jellysmack, a technological unicorn specialized in media.

*I did the Data Analyst training in the bootcamp format. It went very well, the team is very nice and professional. The bootcamp format is quite intense but very well suited to people who are looking for a new job or who are changing careers. It enables you to get a 100% into it! The project is crucial to put into practice what you learn. I also really appreciated the availability and all the workshops organized by the career center, which allows us to be reassured and gives us lots of advice for the job search after the training!*

“



### **Manon HARREWYN**

Assistant Ingénieur recherche & Data Analyst

After a master's degree in marine biology, Manon decided to try the DataScientest adventure in order to add the dual role of Engineer/Data Analyst to her profile. She is now working as an assistant research engineer and data analyst on marine micropollutants.

“ An excellent training! I took the Data Analyst course and I recommend it without moderation. This training allows anyone, beginner or experienced, to acquire solid skills in Data Science that can be used in a wide range of applications. Moreover, the DataScientest team is very attentive and available for every need. Don't hesitate one second more, get involved in Data!

### **Andréas TRUPIN** Data Analyst at Mydral

After studying food engineering, Andréas worked for 5 years in marketing and innovation consulting. He wanted to train as a Data Analyst in order to be able to better support his clients by helping them make the most of their data.



For a professional reconversion, I followed the Data Analyst training of DataScientest in Bootcamp. The training is very well thought out to fit in two and a half months. You have to be 100% dedicated to the training but it's worth it and it also allowed me to take the optional courses and to dig deeper into the topics around the Data Analyst job. The training team was super responsive and the DataScientest teams also set up workshops for recruitment. I would like to thank them for the Recruitment Fair they organized and which allowed me to find a job at the end of my training. A big thank you to them!

“

# TO GO FURTHER

If you wish to strengthen your skills, DataScientest has a variety of training courses on publisher certifications such as Microsoft or AWS that enable you to deepen your knowledge and improve your skills in Data !

## DEEP LEARNING

Are you interested in how Siri works? The creation of autonomous cars? This course is for you! Through two specialization tracks, you will learn all about language processing with Natural Language Processing (NLP), or about image processing with Computer Vision.

## SOFTWARE CERTIFICATION

Following the Microsoft PL-900 certification, you can specialize by preparing for a more technical Microsoft certification.

## POWER BI

Do you want to provide a complete analysis of a dataset and improve your dashboarding skills? This training is for you! Learn to master **Power BI** and get your official Microsoft certification by becoming a **Power BI Data Analyst Associate**.

## EXCEL

Would you like to discover and master **Excel** ? The indispensable technology that is essential for all companies? This course is what you are looking for! Learn to master the different **functionalities** of this tool and thus **develop essential skills** for your (future) job!





**YOU WANT TO BECOME  
A DATA ANALYST ?**

