ZETECH UNIVERSITY

DATA SCIENCE PROGRAMMIG WITH PYTHON

Lesson 1: Introduction to the data science process and the value of learning data science

This is a course for passionately curious that want to work with Data to:

1. Help business leverage data for innovation and success
2. Innovate and predict future trends in business and other industries
3. Learn how to analyze data and provide data- driven insight to make decisions

Learning Objectives

- Basic process of data science
- Python and Jupyter notebooks
- An applied understanding of how to manipulate and analyze uncrated datasets
- Basic statistical analysis and machine learning methods
- How to effectively visualize results

Machine requirements specification for Data Science

- **Internet access**
- **Laptop with following specs**
  - ✓ (Core i 5 upwards, 8 GB RAM, 256 SSD, or 512 HDD & up wards of storage)

Definition of data science.

- **Data science is -** *the field of applying advanced analytics techniques and scientific principles to extract valuable information from data for business decision-making, strategic planning and other uses.*

  - **Data science scope -** *data science will deal with everything, from analyzing complex data, creating new analytics algorithms and tools for data processing and purification, and even building powerful, useful visualizations.*

Some examples of data science are:

- ✓ **Customer Prediction** - *System can be trained based on customer behavior patterns to predict the likelihood of a customer buying a product*
- ✓ **Service Planning** - *Restaurants can predict how many customers will visit on the weekend and plan their food inventory to handle the demand*

Data Terminologies:

Big data - *refers to any large and complex collection of data.*

Data analytics - *is the process of extracting meaningful information from data.*

Data science - *is a multidisciplinary field that aims to produce broader insights.i.e*

- *Scientific methods*
- *Maths and statistics*
- *Programming*
- *Advanced analytics*
- *ML and AI*
- *Deep learning*

Data science terms: *Key differences between these fields*

	Data Analytics	Data Science	Machine Learning
Goal	Extract relevant information from a usually rather small dataset	Conduct operations over various data sources to prove or disprove a certain hypothesis	Develop software that learns by itself by extracting meaning from data
Tools	Involves using analytics applications on structured data	Involves using ML tools to work with both structured and unstructured data	Involves using ML algorithms and analytical models
Scope	Includes predictive modeling, risk analytics, and other	Involves data acquisition, data cleaning, data investigation, etc.	Includes supervised, unsupervised, semi-supervised learning
Output	Trend analysis	Report based on key data	ML model

Key differences between these fields

Data analytics

- you will need to know a programming language, usually R or Python, since these languages have rich libraries that will help you to work with data.
- you will need Structured Query Language (SQL) *to view, manage and access information* you're working with.
- Finally, data analysts often have to present the results of their findings to clients or other stakeholders.
- So you will need to learn how to do data visualization, for example, with the help of *Google Charts, Tableau, Grafana*. You will also need confidence and good presentation skills.

Data science

- A **data scientist -** *is someone who often has to formulate and prove or refute hypotheses.*
  - That is why if you choose this profession, *it's important to have a solid academic background and be able to approach problems systematically and methodically.*
- **Data science** teams often publish papers that report about the results of their experiments and attract public attention to the problems they are working on.
- **Speaking more practically**,
  - you need to *know math and statistics as well as data mining, cleaning, and processing techniques. Knowledge of programming and machine learning techniques* is definitely useful since you often have to build ML models to derive meaning from data.

Machine learning

- **Applied mathematics** is quite an important skill in the arsenal of a machine learning engineer.
  - *As soon as you start working on complex projects, you will discover that out-of-the-box models don't work as well as you would like them to, and you will have to search for solutions.* If you have good knowledge of math theory and statistics, you will be much more efficient at your job.
- **Machine learning specialist is also an engineer,** so *programming is essential.* Python is *the most common choice for machine learning*, however, there are other languages that are gaining popularity in this field such as Julia.

Data science tools & technologies

- This includes programming languages like **R, Python, Julia,** which can be *used to create new algorithms, ML models, AI processes* for big data platforms like *Apache Spark and Apache Hadoop.*
- **Data processing and purification tools** -such as
    - Winpure, Data Ladder.
- **Data visualization tools** -such as
    - Microsoft Power Platform, Google Data Studio, Tableau to **visualization frameworks** like
        - matplotlib and ploty can also be considered as data science tools.
- As **data science** covers everything related to data, any tool or technology that is used in Big Data and Data Analytics can somehow be utilized in the Data Science process.

Why use python in Data science

- Python as a programming language has become very popular in recent times. It has been used in data science, IoT, AI, and other technologies, which has added to its popularity.
- Python is used as a programming language for data science *because it contains costly tools from a mathematical or statistical perspective.*

There are several other reasons why Python is one of the most used programming languages for data science, including:

a. **Speed** - *Python is relatively faster than other programming languages*
b. **Availability** - *There are a significant number of packages available that other users have developed, which can be reused*
c. **Design goal** - *The syntax roles in Python are intuitive and easy to understand, thereby helping in building applications with a readable codebase*

Career opportunity for Data science

1. **Data scientist –** A generalist who knows a bit of everything.
    - ✓ *Deals with all aspects of a project in big companies. Their skill set allow them to overlook a project and guide them from start to finish.*
2. **Data analyst -** Prepare reports that effectively shows the trends and insights gathered from their analysis.
    - ✓ *Data analyst responsible for different tasks such as Visualizing, transforming and manipulating the data. Web analytics tracking and A/B testing analysis.*

3. **Data Engineer -** Responsible for designing, building, and maintaining data pipelines.
    - ✓ *They need to test the ecosystem for the businesses and prepare for data scientist to run their algorithms.*
4. **Data Storyteller -** find the narrative that best describes

Frequently Asked Questions

1. **WHAT JOBS WILL THIS PROGRAM PREPARE ME FOR?**
    - *As a graduate of this program, you will be proficient in the programming skills used in many data analysis and data science roles. Including* Python, R, SQL, Terminal, and Git.

2. **HOW DO I KNOW IF THIS PROGRAM IS RIGHT FOR ME?**
    - *If you are interested in taking the first step into the field of Data Science, this course is for you.* This course will quickly teach you the foundational data programming tools (Python, SQL, Git)
    - **Data Science with Python**, you'll also learn specialized data libraries for Python including *Pandas and Numpy*, and use
        - **Git and the Terminal -** *to share your work and learn about version control*
3. **WHAT IS THE DIFFERENCE BETWEEN THE PYTHON TRACK AND THE R TRACK?**
    - **Python and R Both -** *tracks cover the same fundamental concepts, use data sets but use a different programming language.*
        - The SQL, command line, and Git - *curriculum is the same in both tracks*

Target groups: suitable for Learning data science.

- Statisticians • IT Professionals • ICT Professionals • data managers • Software Developers and Architects, • Business Intelligence Professionals • Project Managers, • Engineers , Bio data specialist ,Aspiring Data Scientists, • University students looking to begin a career in Big Data Analytics etc…

Anaconda IDE for Python Data science.

- **Anaconda is a distribution of Python and R programming languages for scientific computing** (data science, machine learning, large scale data processing, predictive modelling) ***its Aims to simplify package management and deployment.***

- *How to Install Anaconda!* *Video. Link.*

  https://www.youtube.com/watch?v=T8wK5loXkXg

Installing Anaconda on Windows Tutorial

https://www.datacamp.com/tutorial/installing-anaconda-windows

1. **What is Data Science? | Introduction to Data Science | Data Science for Beginners | Simplilearn**

https://www.youtube.com/watch?v=KxryzSO1Fjs

2. **What REALLY is Data Science? Told by a Data Scientist**
   https://www.youtube.com/watch?v=xC-c7E5PK0Y

END of INTRODUCTION PART!!!!!