

INTRODUCTION TO DATA SCIENCE

Course Description

Author: Eng. Carlos Andrés Sierra, M.Sc.
carlos.andres.sierra.v@gmail.com

Lecturer
Computer Engineer
School of Engineering
Universidad Distrital Francisco José de Caldas

2024-II



Outline

1 You don't know who I am

2 Course Overview

3 Grading & Rules

4 Bibliography



Outline

- 1 You don't know who I am

- ## 2 Course Overview

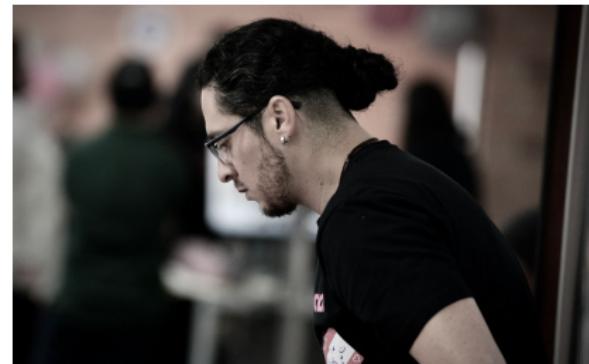
- ### ③ Grading & Rules

- ## 4 Bibliography



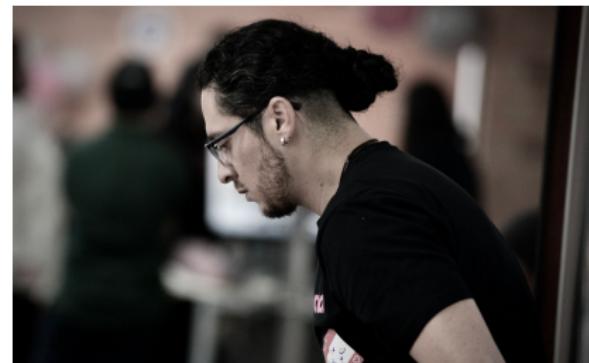
Academic Experience

- **Computer Engineer, M.Sc. in Computer Engineering, researcher** for 15 years.
- 7 years as **full-time associate professor** at colleges, for Computer Engineering programs.
- 3 years as **lecturer professor** for both colleges and government STEM programs.
- **Speaker** in Colombia, Brasil, Bolivia, at IEEE events and colleges.



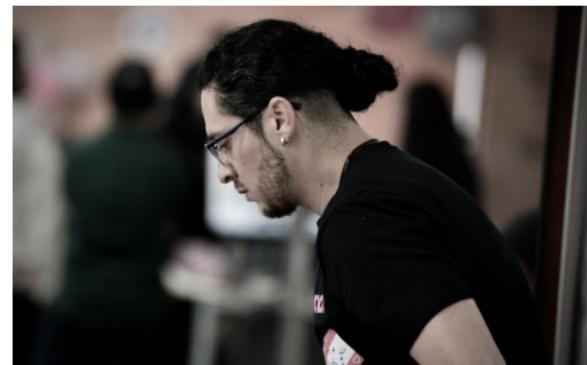
Academic Experience

- **Computer Engineer, M.Sc. in Computer Engineering, researcher** for 15 years.
- 7 years as **full-time associate professor** at colleges, for **Computer Engineering programs**.
- 3 years as **lecturer professor** for both colleges and government **STEM** programs.
- **Speaker** in Colombia, Brasil, Bolivia, at **IEEE** events and colleges.



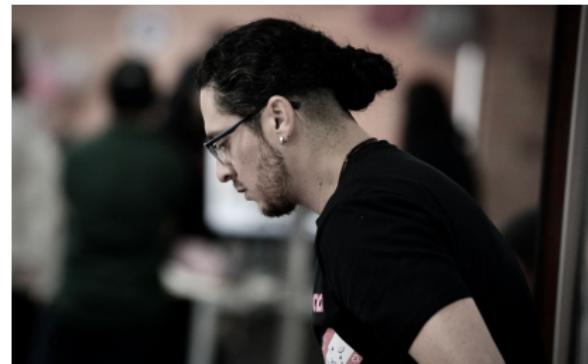
Academic Experience

- **Computer Engineer, M.Sc. in Computer Engineering, researcher** for 15 years.
- 7 years as **full-time associate professor** at colleges, for **Computer Engineering programs**.
- 3 years as **lecturer professor** for both colleges and government **STEM programs**.
- Speaker in Colombia, Brasil, Bolivia, at **IEEE events and colleges**.



Academic Experience

- **Computer Engineer, M.Sc. in Computer Engineering, researcher** for 15 years.
- 7 years as **full-time associate professor** at colleges, for **Computer Engineering programs**.
- 3 years as **lecturer professor** for both colleges and government **STEM programs**.
- **Speaker** in Colombia, Brasil, Bolivia, at **IEEE** events and colleges.



Non-academic Experience



- PyCon Colombia and Python Bogotá **co-organizer**. Collaborations in *ScipyLATAM* and *Jupyter LATAM*.
- 3 years as **software engineer** for several companies in Colombia.
- 3 years as **Technical Leader of Machine Learning and Data Science** in a USA startup.
- 1 year as **MLOps Engineer** for a Fintech in LATAM.



Non-academic Experience



- PyCon Colombia and Python Bogotá **co-organizer**. Collaborations in *ScipyLATAM* and *Jupyter LATAM*.
- 3 years as **software engineer** for several companies in Colombia.
- 3 years as **Technical Leader** of Machine Learning and Data Science in a USA startup.
- 1 year as **MLOps Engineer** for a Fintech in LATAM.



Non-academic Experience



- PyCon Colombia and Python Bogotá **co-organizer**. Collaborations in *ScipyLATAM* and *Jupyter LATAM*.
- 3 years as **software engineer** for several companies in Colombia.
- 3 years as **Technical Leader** of Machine Learning and Data Science in a USA startup.
- 1 year as MLOps Engineer for a Fintech in LATAM.



Non-academic Experience



- PyCon Colombia and Python Bogotá **co-organizer**. Collaborations in *ScipyLATAM* and *Jupyter LATAM*.
- 3 years as **software engineer** for several companies in Colombia.
- 3 years as **Technical Leader** of Machine Learning and Data Science in a USA startup.
- 1 year as **MLOps Engineer** for a Fintech in LATAM.



Outline

1 You don't know who I am

2 Course Overview

3 Grading & Rules

4 Bibliography



Overview

This course is designed to introduce undergraduate students to **foundations** of **data science**, **data engineering**, and **database systems**. Also, this course will provide you with the necessary **tools** to **design**, **implement**, and **maintain** **data science systems** for software applications. In particular, this course will focus on **python technologies** as **main tools**.

Classes will consist of **lectures**, **discussions**, practical examples, and **workshops**. Also, you must take some readings from *data science*. In addition, there will be a **course-long project**, as well **one course test**, and **four workshops**.



Overview

This course is designed to introduce undergraduate students to **foundations** of **data science**, **data engineering**, and **database systems**. Also, this course will provide you with the necessary **tools** to **design**, **implement**, and **Maintain** **data science systems** for software applications. In particular, this course will focus on **python technologies** as **main tools**.

Classes will consist of **lectures**, **discussions**, practical examples, and **workshops**. Also, you must take some readings from *data science*. In addition, there will be a **course-long project**, as well **one course test**, and **four workshops**.



Goals

The main goal of this course is to provide undergraduate students with different ~~models~~, ~~concepts~~ and ~~tools~~ for solving problems using data science in order to provide software solutions with a good level of quality.

At the end of this course you should be able to create a full software data science solution with a good level of quality metrics. Also, you should be able to design robust data science systems using state-of-the-art tools.



Goals

The **main goal** of this course is to provide undergraduate students with different **models**, **concepts**, and **tools** for **solving problems** using **data science** in order to provide **software solutions** with a good level of **quality**.

At the end of this course you should be able to **create** a **full software data science solution** with a good level of **quality** metrics. Also, you should be able to **design** robust **data science systems** using **state-of-the-art tools**.



Prerequisites

This is a basic course, so you must have some knowledge in:

- **Programming** in Python, C++, Java, Golang. **I am kidding!**
- Programming basics (variables, loops, functions, etc.).
- Mathematics basics (calculus, algebra, statistics).
- Git basic usage, and GitHub basic usage.
- Use of IDEs like VS Code, Eclipse, or PyCharm.
- Mandatory: Desire to learn and passion for data science.
- Recommended: Linux OS, Containers, Jupyter Notebooks, DataBases, and Agile Methodologies.



Prerequisites

This is a basic course, so you must have some knowledge in:

- **Programming** in Python, C++, Java, Golang. **I am kidding!**
- **Programming basics** (variables, loops, functions, etc.).
- **Mathematics basics** (calculus, algebra, statistics).
- **Git** basic usage, and **GitHub** basic usage.
- Use of **IDEs** like **VS Code**, Eclipse, or PyCharm.
- Mandatory: **Desire to learn** and **passion for data science**.
- Recommended: Linux OS, Containers, Jupyter Notebooks, DataBases, and Agile Methodologies.



Prerequisites

This is a basic course, so you must have some knowledge in:

- **Programming** in Python, C++, Java, Golang. **I am kidding!**
- **Programming basics** (variables, loops, functions, etc.).
- **Mathematics** basics (calculus, algebra, statistics).
- **Git** basic usage, and **GitHub** basic usage.
- Use of **IDEs** like **VS Code**, Eclipse, or PyCharm.
- Mandatory: **Desire to learn** and **passion for data science**.
- Recommended: Linux OS, Containers, Jupyter Notebooks, DataBases, and Agile Methodologies.



Prerequisites

This is a basic course, so you must have some knowledge in:

- **Programming** in Python, C++, Java, Golang. I am kidding!
- **Programming basics** (variables, loops, functions, etc.).
- **Mathematics** basics (calculus, algebra, statistics).
- **Git** basic usage, and **GitHub** basic usage.
- Use of IDEs like VS Code, Eclipse, or PyCharm.
- Mandatory: Desire to learn and passion for data science.
- Recommended: Linux OS, Containers, Jupyter Notebooks, DataBases, and Agile Methodologies.

PLttty



Prerequisites

This is a basic course, so you must have some knowledge in:

- **Programming** in Python, C++, Java, Golang. **I am kidding!**
- **Programming basics** (variables, loops, functions, etc.).
- **Mathematics** basics (calculus, algebra, statistics).
- **Git** basic usage, and **GitHub basic usage**.
- Use of **IDEs** like **VS Code**, Eclipse, or PyCharm.
- Mandatory: **Desire to learn** and **passion** for **data science**.
- Recommended: Linux OS, Containers, Jupyter Notebooks, DataBases, and Agile Methodologies.



Prerequisites

This is a basic course, so you must have some knowledge in:

- **Programming** in Python, C++, Java, Golang. **I am kidding!**
- **Programming basics** (variables, loops, functions, etc.).
- **Mathematics** basics (calculus, algebra, statistics).
- **Git** basic usage, and **GitHub basic usage**.
- Use of **IDEs** like **VS Code**, Eclipse, or PyCharm.
- Mandatory: **Desire to learn** and **passion** for **data science**.
- Recommended: Linux OS, Containers, Jupyter Notebooks, DataBases, and Agile Methodologies.



Prerequisites

This is a basic course, so you must have some knowledge in:

- **Programming** in Python, C++, Java, Golang. I am kidding!
- **Programming basics** (variables, loops, functions, etc.).
- **Mathematics** basics (calculus, algebra, statistics).
- **Git** basic usage, and **GitHub** basic usage.
- Use of **IDEs** like **VS Code**, Eclipse, or PyCharm.
- Mandatory: **Desire to learn** and **passion** for **data science**.
- Recommended: **Linux OS**, **Containers**, **Jupyter Notebooks**, **DataBases**, and **Agile Methodologies**.

alongside

colab



Syllabus I

Period	Topic	Time
Period I	Data Science Basic Concepts	2 sessions
	Python Introduction	2 sessions
	Data Manipulation with NumPy	2 sessions
	Data Manipulation with Pandas	4 sessions
	OpenData, ETLs, EDA	4 sessions
Period II	Information Visualization	2 sessions
	Charting with MatPlotLib	4 sessions
	Charting with Seaborn	2 sessions
Period II	Fundamentals of Machine Learning	2 sessions
	Supervised Learning and Scikit-Learn	4 sessions
	Models Evaluation and MLOps	4 sessions

Table: Schedule for Course



Outline

1 You don't know who I am

2 Course Overview

3 Grading & Rules

4 Bibliography



Grades Percentages

Period	Item	Percentage
Period I	Workshop 1	10%
	Workshop 2	10%
Period I	Workshop 3	10%
	Workshop 4	10%
Period III	Paper + Poster	15%
	Final Test	30%
	Course Project	15%

Table: Course Grades Distribution



Don't hate the player, hate the game

- All assignments must be submitted on time and in english. Grammar and spelling will not be evaluated.
- Copying and pasting from internet is forbidden. Please, develop your own solutions.
- Class attendance is not mandatory. If you miss classes, you must study by yourself. If you abandon the course, you will fail.
- No cell-phones, no smartwatches, no whatsapp, no tinder, no smartanything. Just you and your brain. Pay attention at clase.
- Communications with me must be done by institutional email or by slack. I will not answer any question by WhatsApp.



Don't hate the player, hate the game

- All assignments must be submitted on time and in english. Grammar and spelling will not be evaluated.
- Copying and pasting from internet is forbidden. Please, develop your own solutions.
- Class attendance is not mandatory. If you miss classes, you must study by yourself. If you abandon the course, you will fail.
- No cell-phones, no smartwatches, no whatsapp, no tinder, no smartanything. Just you and your brain. Pay attention at clase.
- Communications with me must be done by institutional email or by slack. I will not answer any question by WhatsApp.



Don't hate the player, hate the game

- All assignments must be submitted on time and in english. Grammar and spelling will not be evaluated.
- Copying and pasting from internet is forbidden. Please, develop your own solutions.
- Class attendance is not mandatory. If you miss classes, you must study by yourself. If you abandon the course, you will fail.
- No cell-phones, no smartwatches, no whatsapp, no tinder, no smartanything. Just you and your brain. Pay attention at clase.
- Communications with me must be done by institutional email or by slack. I will not answer any question by WhatsApp.



Don't hate the player, hate the game

- All assignments must be submitted on time and in english. Grammar and spelling will not be evaluated.
- Copying and pasting from internet is forbidden. Please, develop your own solutions.
- Class attendance is not mandatory. If you miss classes, you must study by yourself. If you abandon the course, you will fail.
- No cell phones, no smartwatches, no whatsapp, no tinder, no smartanything. Just you and your brain. Pay attention at clase.
- Communications with me must be done by institutional email or by slack. I will not answer any question by WhatsApp.



Don't hate the player, hate the game

- All assignments must be submitted on time and in english. Grammar and spelling will not be evaluated.
- Copying and pasting from internet is forbidden. Please, develop your own solutions.
- Class attendance is not mandatory. If you miss classes, you must study by yourself. If you abandon the course, you will fail.
- No cell-phones, no smartwatches, no whatsapp, no tinder, no smartanything. Just you and your brain. Pay attention at clase.
- Communications with me must be done by institutional email or by slack. I will not answer any question by WhatsApp.

Caviguez@udistrital.edu



Code of Conduct

- Always be **respectful** to your **classmates and to me**. You must be **kind** with **everyone** inside (*and outside*) the classroom.
- There is **no** a better programming language, tool, or technology. There are only **better** or **worse** solutions.
- You must be **honest** with your work. If you don't know something, just **ask** me. I will be **happy** to help you.
- You must be **responsible** with your work. If you don't submit **on time**, please **don't cry**.
- You must **not be annoying**, or affect the **classroom environment**. If you do, I will **ask you to leave** the classroom.



Code of Conduct

- Always be **respectful** to your **classmates and to me**. You must be **kind** with **everyone** inside (*and outside*) the classroom.
- There is **no a better programming language**, tool, or technology. There are only **better** or **worse** solutions.
- You must be **honest** with **your work**. If you don't know something, just **ask me**. I will be **happy to help you**.
- You must be **responsible** with **your work**. If you don't submit **on time**, please **don't cry**.
- You must **not be annoying**, or affect the **classroom environment**. If you do, I will **ask you to leave the classroom**.



Code of Conduct

- Always be **respectful** to your **classmates and to me**. You must be **kind** with **everyone** inside (*and outside*) the classroom.
- There is **no a better programming language**, tool, or technology. There are only **better** or **worse** solutions.
- You must be **honest** with **your work**. If you don't know something, just **ask** me. I will be **happy to help you**.
- You must be **responsible** with **your work**. If you don't submit **on time**, please **don't cry**.
- You must **not be annoying**, or affect the **classroom environment**. If you do, I will **ask you to leave** the classroom.



Code of Conduct

- Always be **respectful** to your **classmates and to me**. You must be **kind** with **everyone** inside (*and outside*) the classroom.
- There is **no a better programming language**, tool, or technology. There are only **better** or **worse** solutions.
- You must be **honest** with **your work**. If you don't know something, just **ask** me. I will be **happy to help you**.
- You must be **responsible** with **your work**. If you don't submit **on time**, please **don't cry**.
- You must **not be annoying**, or affect the **classroom environment**. If you do, I will **ask you to leave** the classroom.



Code of Conduct

- Always be **respectful** to your **classmates and to me**. You must be **kind** with **everyone** inside (*and outside*) the classroom.
- There is **no a better programming language**, tool, or technology. There are only **better** or **worse** solutions.
- You must be **honest** with **your work**. If you don't know something, just **ask** me. I will be **happy to help you**.
- You must be **responsible** with **your work**. If you don't submit **on time**, please **don't cry**.
- You must **not be annoying**, or affect the **classroom environment**. If you do, I will **ask you** to **leave** the classroom.



Outline

1 You don't know who I am

2 Course Overview

3 Grading & Rules

4 Bibliography



Bibliography

Recommended bibliography:

- **Data Science for Business**, by Foster Provost and Tom Fawcett.
- **Python for Data Analysis**, by Wes McKinney.
- **Data Science from Scratch**, by Joel Grus.
- **Python Data Science Handbook**, by Jake VanderPlas.
- **Introduction to Machine Learning with Python**, by Andreas C. Müller and Sarah Guido.
- **Machine Learning Yearning**, by Andrew Ng.
- **Effective Pandas 2** by Matt Harrison.



Outline

1 You don't know who I am

2 Course Overview

3 Grading & Rules

4 Bibliography



Thanks!

Questions?



www.linkedin.com/in/casierrav

