Intro to Data Science

Group of Horribly Optimistic STatisticians

Maksymilian Norkiewicz, Jędrzej Ogrodowski

Who are we?

We are second-year students majoring in computer science and artificial intelligence.

We both started to be interested in the area related to data analysis and we created several small projects together. Now, we want to give you the basics and tools that we have learned.

For communication:

Discord

Messenger

Attendance:

You can record your attendance in an Excel file which you can find on Discord

What is Data Science?

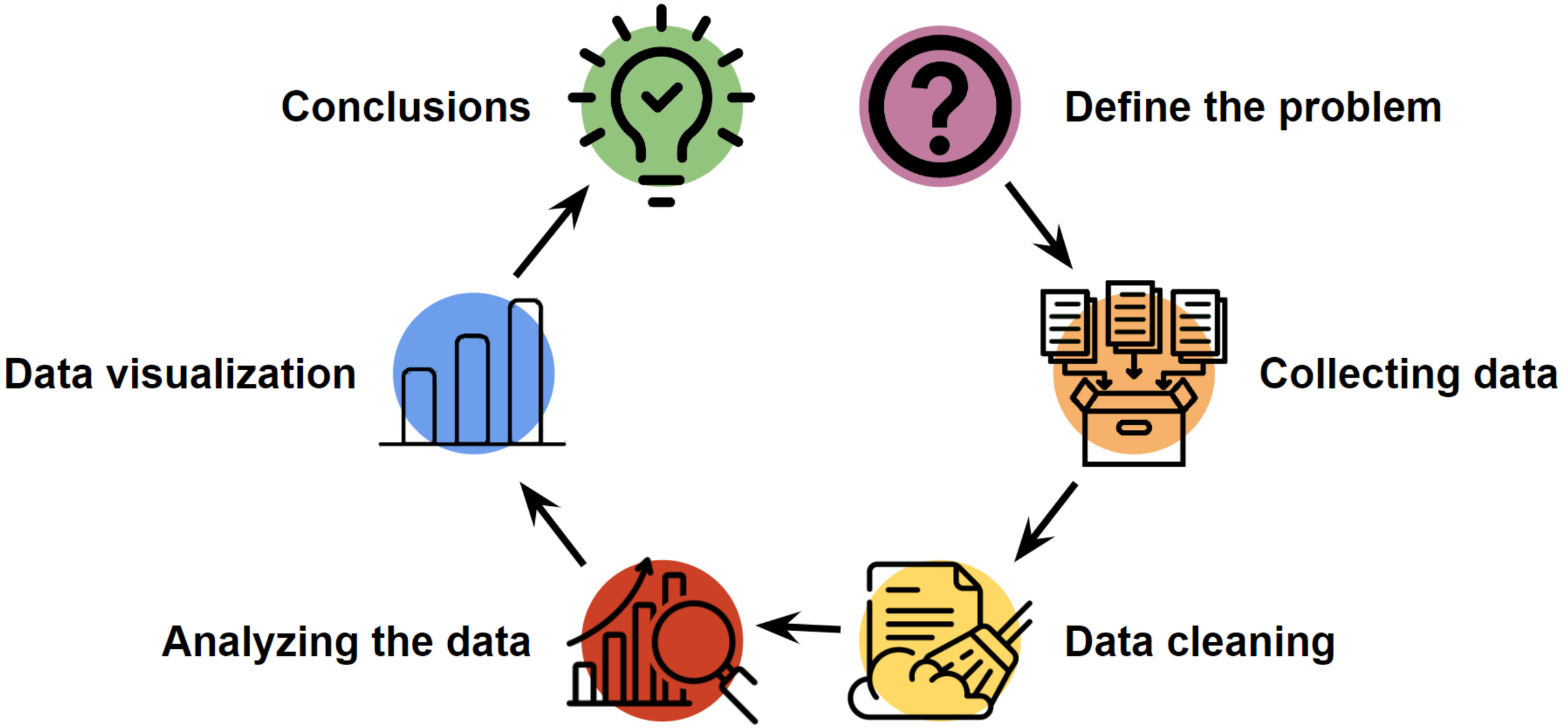
Data Science is science field that uses statistics, computing, data processing, algorithms and visualizations to extract knowledge from given data.

We try to control the chaos in the data and present it in the best possible form and draw conclusions.

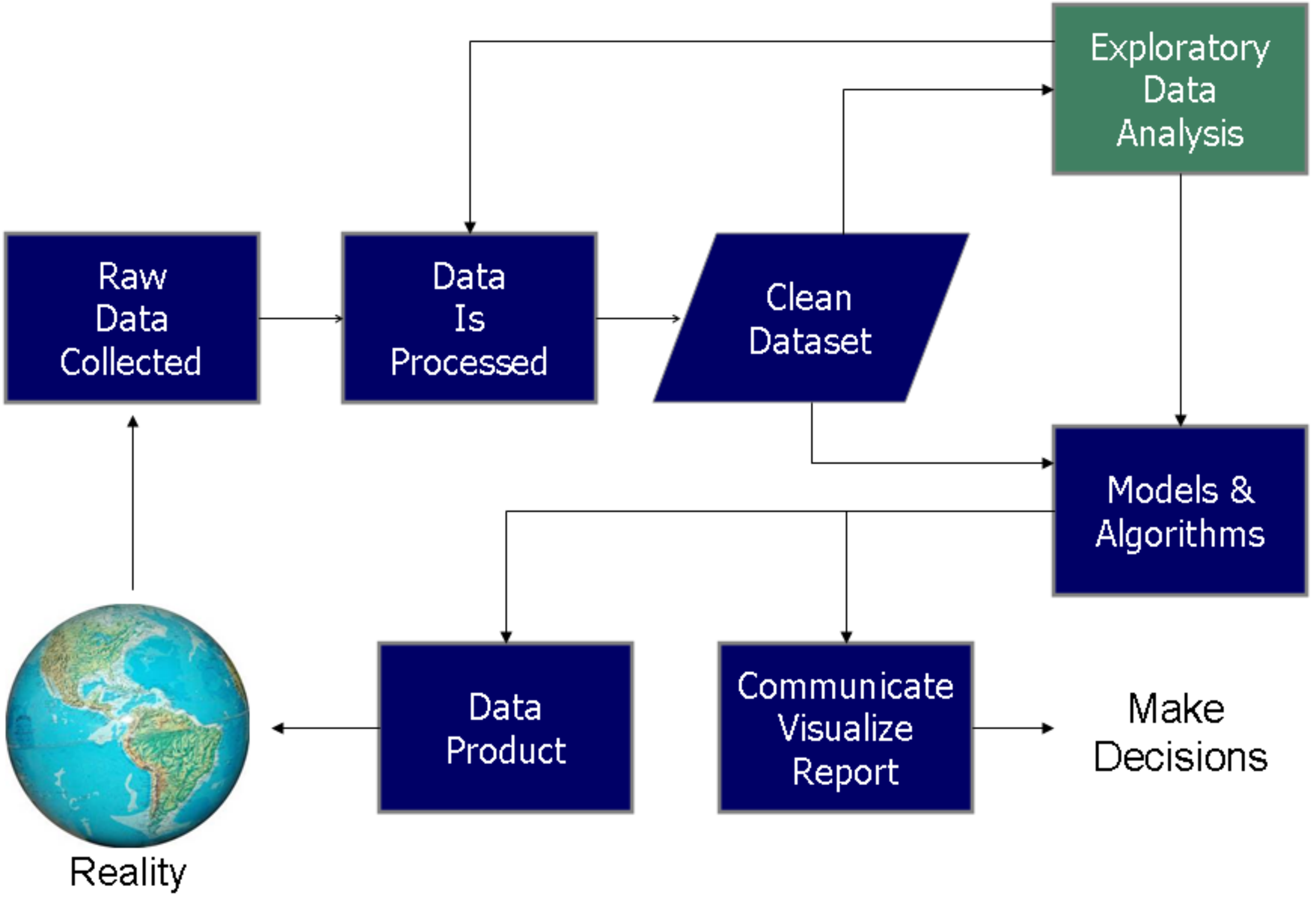
Imagine. Every time you walk through the city, you see food delivery guys going crazy on their bikes. After quick thought, you realise that they are basically everywhere. Now we think, if it is so popular, how do people rate each of these applications?

Let’s create a plan.

- Define a problem.
- Gather data.
- Clean data.
- Analyse data.
- Visualize it.
- Conclusions.


Conclusions
Define the problem
Data visualization
Collecting data
Analyzing the data
Data cleaning

Data Science Process



Who is this section for?

The section is intended for beginners who want to find themselves at the beginning of their path in data science. We will discuss simple topics, reducing patterns to a minimum and focusing on tools.

What’s the plan?

- NumPy and Linear Algebra.
- Pandas and data manipulation.
- Statistics with examples in SciPy.
- Data visualization with Matplotlib.
- Exploratory data analysis.
- Project.

At the end of our journey, we will want to create a project summarizing the acquired knowledge. Our plan is to clean the data, prepare a data set, visualize it, sample it, calculate some stats and use a basic ML algorithm.

What after that?

- You can collect data.
- You can transform this data into a form to work with.
- You can explore data.
- You can visualise it.
- You can draw conclusions.
- **You got solid basics for other sections at GHOST.**

Handouts

All handouts, presentations, links, projects, exercises are on ours GitHub accounts.

Repository.

Bibliography

- W. McKinney, Python for Data Analysis
- Claus O. Wilke, Fundamentals of Data Visualization
- StatSoft Electronic Statistics Book
- P. Bruce, A.Bruce, P Gedeck, Statystyka praktyczna w data science
- C. N. Knaflic, Storytelling with Data

Let’s start!