

# DATA MINING

## Web scraping project

Scrapping a car website (wandaloo)

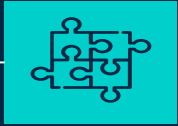
Realized by :

**ISBAINE** Mohamed & **LABRIJI** Saad

Supervised by :

**EL ASRI** Ikram

# TABLE OF CONTENTS



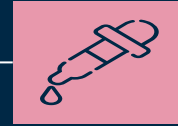
01

Introduction & problem statement



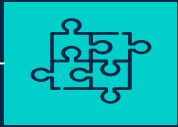
02

Divide Long-Term Goals into Smaller Tasks



03

Technologies & tools used in this project



04

Collecting & preprocessing the data



05

Choosing the machine learning models adapted to the dataset



06

Conclusion

# Introduction & problem statement

01

# Introduction & problem statement

**Pricing a car** can usually be a **difficult task** even for car enthusiasts and experts because **cars have many features that affect prices more or less strongly**, which could lead people without any knowledge in vehicles to be easily ripped off. To solve this kind of problem, we thought my teammate and I that we could use the modest knowledge we acquired in data science courses this year, to develop an machine learning model to price cars more accurately, or at least get an idea of the price range.



Divide **Long-Term** Goals  
into **Smaller** Tasks

02

# Divide Long-Term Goals into Smaller Tasks

Find a **well-structured website** that contains **correct** and **structured** car **data**

STEP 01

STEP 02

Scraping, Processing and **Cleaning** the web data to create a **useful dataset**.

**Evaluation** and **deployment** of **models**, verification of results & **prediction of certain car prices**.

STEP 03

STEP 04

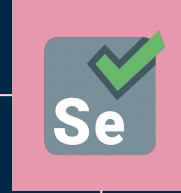
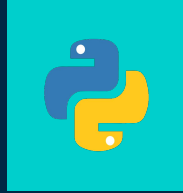
**Concluding** & **what's next?**

## Technologies & tools used in this project

03

# Technologies & tools used in this project ?

Python

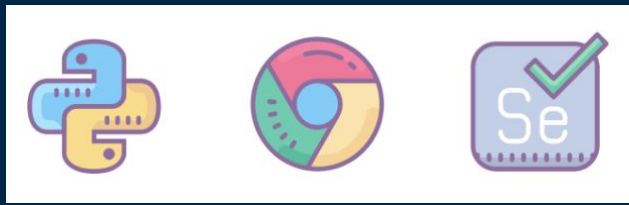


Selenium



Chrome





# WHAT IS SELENIUM?

**Selenium** Python is an **open-source web-based automation tool** which provides a **simple API** to write **functional or acceptance tests using Selenium WebDriver**.

## WHY IT IS USED?

**Selenium** is an open-source tool that **automates web browsers**. It provides a single interface that **lets us write test scripts in programming languages** like Ruby, Java, NodeJS, PHP, Perl, Python, and C#, among others.

## WHAT ARE SOME ADVANTAGES OF SELENIUM?

- **Selenium** is an **Open Source Software**.
- **Selenium** supports **various programming languages** to write programs (Test scripts)
- **Selenium** uses **fewer Hardware resources**.
- **Selenium** supports **various Browsers**.
- **Selenium** supports **Parallel Test Execution**.

# Collecting & preprocessing the data

04

Choosing the **machine  
learning models** adapted  
to the dataset

05

# Conclusion

06



# THANKS !

Realized by :

**ISBAINE** Mohamed & **LABRIJI** Saad

Supervised by :

**EL ASRI** Ikram