# Youssef DIR

Recently graduated with a Master's degree **in Data & Business Intelligence** (ranked 6th nationally) from the University of Tours, I am seeking  **Data scientist - Data engineer** post**,** available starting mars 2025.

✉ | youssefdir37@gmail.com |📱  0666936752  | 🔘 Dirtoussef  | 🔗 Youssef Dir ↗  h t t p s : / / y o u s s e f d i r . n e t l i f y . a p p /  |

## Education

| 2022-2024 | **Université of Tours** , **M2 Business, Data intelligence :** Big Data,  Data Mining , Maching learning , Web scraping , Exploratory Data Analysis  ( Python , Rstudio ) , Microeconometrics , APIS , Pricing. SQL Data Management. | *Tours, France* |
|---|---|---|
| 2018-2022 | **Université of Tours** , **M2 Business, Data intelligence :** Econometrics,  Statistics , Probability , R , Python , Microeconomics , Excel (VBA ) ,HTML - CSS , Macroeconomics , Mathematiques. | *Tours, France* |

## Experiences

**AVANCI MVGROUP**                                                                                                           *Tours, France*
**Data scienitst - Intership (6 mois)**                                                                         *Mai 2024 - Oct 2024*

**Mission 1** : **Creation and Management of a Simulated Database - SQL**
- Generation of realistic data using the Faker library (**Python**) to simulate a dataset tailored for business analyses.
- Implementation of data on SQL Server, managing relationships and structuring tables for efficient access.
- Data cleaning: handling missing values, detecting inconsistencies, and preparing data for exploratory analysis.

**Mission 2 : Development of Dynamic Dashboards and Data Visualization with Streamlit - Python**
- Integration of visualization tools such as **Plotly** and **Matplotlib** to present insights (segmentation, business performance).
- Creation of advanced visualizations: interactive map to analyze regional audience, charts for customer segmentation, and performance tracking.
- Implementation of a dedicated tab for data quality visualization.

**Mission 3 : Analyse prédictive et modélisation avancée**
- Statistical analysis and data cleaning (handling missing values, detecting anomalies).
- Selection of key features (top features) by evaluating their relative importance in predictions.
- Statistical analysis and data cleaning (handling missing values, detecting anomalies).
- Redundancy detection with **PCA** between variables to optimize the selection of relevant features.
- Implementation of a sampling strategy to balance the dataset in case of imbalanced classes.
- Training of Random Forest and Boosting (XGBoost) models to predict target variables related to customer behaviors and characteristics.
- Evaluation of model performance using key metrics such as accuracy and **AUC ROC**.

**PROBTP  Finance**                                                                                                        *Paris, France*
**Data analyst Finance Intership (3 month)**                                                      *Juin 2022 - Sep 2023*
- Development of a **VBA** script to automate the reading of PDF files, extract key information, and organize it into structured tables.
- Development of a VBA-based tool to track SCPI (Société Civile de Placement Immobilier values, including purchase values, Realization values,subscription prices, etc.

## Project

**Academic Projects**                                                                                                      *Tours, France*
**Data Scraping and Car Price Prediction**                                                                  *2023/2024*
- Data extraction through web scraping on AutoScout (used car sales) using BeautifulSoup.
- Nettoyage, Data cleaning, processing, and analysis with **Pandas** and **NumPy**.
- Modeling and prediction of prices using Machine Learning algorithms (**Boosting, Random Forest**, **K-Means**).
- Documentation of the project in **RMarkdown** for a structured and clear presentation.

**RShiny Application for Text Analysis and Interactive Visualization**
- Data cleaning, processing, and analysis.
- Sentiment analysis of users by country.
- Statistics and interactive visualization table of results.

## Skilss

| | | |
|---|---|---|
| ♟ | **Langues** | **French** (Bilingual), **Anglais** (B2) |
| ☰ | **Langages** | Python , Rstudio, PowerBI ,Dataiku, Shell |
| ⚙ | **Data science** | Boosting (Xgboost), Random Forest,NLP, Régession linéaire, Clustering (**Sklearn – Tidymodels**) |
| >_ | **Data Apps** | Streamlit -  Rshiny |
| ⬓ | **Base de données** | SQL Server , My SQL , PostgreSQL |
| ↑↑ | **Devops** | Gitlab, Github |
| ☁ | **Cloud Data** | AWS, Snowflake |
| 🗎 | **Reportings** | Quarto ,Rmarkdown, |