

# Mustapha Bouhsen

Canadian citizen

[Website](#)

Montreal (Québec)

514 603-0115

[bouhsen.m@gmail.com](mailto:bouhsen.m@gmail.com)

[www.linkedin/in/mustapha-bouhsen](https://www.linkedin/in/mustapha-bouhsen)

Langue: French, English, Arabic

## SKILLS

---

Machine Learning, Deep Learning, Data Engineering, Data Analysis, Data Visualization

Probability, Statistic, Risk Theory, Survival analysis, Actuarial mathematic

Languages: R, Python, SQL, C++, Git, Bash

Tech Tools: Databricks, Microsoft Azure, Synapse, Tableau, Power BI

Frameworks: PySpark, PyTorch, Pandas, NumPy, Matplotlib, Sci-Kit Learn, ggplot.

## EDUCATION

---

**M.Sc. in Data Science**

HEC Montréal

2022 - 2023

**B.Sc. in Actuarial Science.**

École d'actuariat - Université Laval

2019 - 2022

## EXPERIENCE

---

**Data Engineer intern at Bombardier**

2023/09-

- Create **PySpark** ETL pipeline to transfer files in parquet format to **Microsoft Azure storage**.
- Created **ML algorithms** to predict plant part prices.
- Implemented **pipelines** and **SQL** scripts for tracking and communicating automated migration progress via email.

Achievements:

- Optimize **ETLs** that initially took at least **48 hours** to reduce it to just **8 minutes (99.72%)**, improving business efficiency and productivity.
- Developed a robust general-purpose library using **Apache Spark** to streamline and accelerate the creation of data processing pipelines.
- Refactor the transformation process from using **Pandas** to **PySpark** for process optimization.
- Convert **SQL** to **PySpark** code for enhanced flexibility and leverage Spark's distributed computing capabilities, contributing to optimized data workflows.

**Data Scientist searcher at Laval University**

2022/04-2022/09

- Employ **Pandas** and **Scikit-learn** for thorough data cleaning, addressing missing values, and handling outliers.
- Conduct **statistical inference**, employing Regression Analysis and Hypothesis Testing, to gain insights into the influence of variables on the overall condition of golf courses.
- Implement **machine learning** algorithms, including **Random Forest**, **KNN**, and **Gradient Boosting**, utilizing both **Python** and **R** environments. These models predict the impact of various chemical components on the condition of golf courses.

**Data Scientist searcher at Laval University**

2021/04-2021/09

- Address missing data and perform data cleaning using **R** and **python**.
- Run optimization algorithms on Compute Canada using **C++**.
- Automate tasks using **R** and **Python** and reduce processing time that **takes hours** in just **one click**.

## Projects

---

- [Financial Data Intelligence Suite](#)
- [A STUDY OF BANK CUSTOMER CHURN](#)
- To see all my projects, please visit my [website](#)