



Project 2 - Shark Attacks

Data Cleaning and Manipulation

Main Objectives

The dataset we'll provide you contains significantly messy data. Your job is to apply the different cleaning and manipulation techniques to generate a cleaner CSV version of this data.



Main Objectives - Make yourself focused

Answer a 'business' question!

Look for something you want to answer about shark attacks and try to answer that using data. This will focus your cleaning and data manipulation through the process.



The Dataset

- Go to [kaggle.com](https://www.kaggle.com) and create an account
- Go to the search bar and look for 'Global Shark Attacks'
- Download the dataset
- For more info: <https://www.sharkattackfile.net/>



Hints

- Use their website to understand the problem. Understanding the problem is a crucial step to developing good quality data solutions.
- Be focused. The dataset is really messy. Every clean you make should have the goal to contribute to your answer.



Deliverables

- A README.md file on GitHub describing your objectives and results.
- The link to the Jupyter notebook (or the GitHub project)

MORE Objectives

- One of the objectives of the current project is to show you the need of some techniques we will see in the future in the course. Of course, you don't need to use them for now.
- In the process of answering your question, we believe you'll feel the need of some tools like:
 - Visualization techniques
 - Statistics
 - More data sources

Deadline

