

# Struktura folderów

Cały projekt zawarty jest w folderze o nazwie **"Earthquake Data"**.

Folder ten zawiera 4 następujące foldery:

1. **Analysis Data**

Folder ten zawiera oczyszczone oraz przetworzone pliki danych.

2. **Command Files**

W tym folderze znajdują się pliki, które zawierają kody do przetwarzania i analizy danych.

3. **Documents**

Folder ten zawiera kopie pracy końcowej, Data Appendix oraz plik Readme

4. **Original Data**

Folder, który zawiera oryginalne dane oraz folder z Metadanymi

## Ścieżka analizy danych.

1. Importowanie pliku z oryginalnymi danymi, które znajdują się w pliku Original Data, znajduje się w skrypcie "data\_work\_earthquake.ipynb" w folderze Command Files.
2. Przetwarzanie danych: Przetwarzanie danych znajduje się w pliku Analiza\_danych.ipynb w folderze Command Files.
  - Na samym początku zostały zmienione nazwy column na krótsze i bardziej ogólne.
  - Następnie w kolumnie, która dotyczyła pieniędzy, wartości NaN zostały zmienione na "Prefer not to answer".
  - W kolejnym kroku wszystkie wiersze, w których występował jakikolwiek NaN zostały usunięte. Wszystkie odpowiedzi były obiektywne, dlatego nie dało się ich zastąpić żadnymi innymi danymi.
  - Na koniec zostały zaktualizowane wszystkie indeksy danych.
  - Przetworzone dane zostały zapisane w folderze Analysis Data jako "earthquake\_data\_clean.csv".
3. Z przetworzonych danych w tym samym pliku "data\_work\_earthquake.ipynb" została utworzona tabela, która łączy wiek, płeć i odpowiedź na pytanie "Do you think the "Big One" will occur in your lifetime?". Tabela ta jako plik csv została zapisana w folderze Analysis Data jako "earthquake\_data\_newframe.csv"

4. W kolejnym kroku został utworzony plik Data Appendix w folderze Documents, który zawiera książkę kodów i przewodnik użytkownika dla utworzonych plików do analizy danych. Tabelę i wykresy wykorzystane do tworzenia tego dodatku zostały wygenerowane z pliku "data\_appendix\_earthquake.ipynb" z folderu Command Files oraz zapisane w pliku Analysis Data. Na ich podstawie w pliku Data Appendix zostały opisane wszystkie zmienne potrzebne do analizy danych.
5. W tym punkcie przechodzimy już do zasadniczej analizy danych. Aby wygenerować interesujące nas rezultaty, należy uruchomić skrypt "the\_results\_earthquake.ipynb", który znajduje się w folderze Command Files. Znajdują się tam instrukcje do rysowania wykresów, które zawierają interesujące nad wyniki.
6. Całość projektu w formie papierowej została umieszczona w pliku Documents pod nazwą "Final\_paper.pdf"

## PODSUMOWANIE

W celu uzyskania powyższych rezultatów należy odpalić po kolei skrypty:  
"data\_work\_earthquake.ipynb" -> "data\_appendix\_earthquake.ipynb" ->  
"the\_results\_earthquake.ipynb"

## UWAGA DO PROJEKTU

- W powyższych skryptach użyta jest lokalna ścieżka do otwierania plików z projektu. Należy zmienić ją, aby dane wczytywały się poprawnie.
- W projekcie użyty był Python wersja 3.8.3 oraz rozszerzenia Pandas, matplotlib.pyplot, seaborn