

## Questions/Problem statement

- o List of Questions (Fake News).
  - How many fake and real news in the dataset?
  - What are the languages of news sites?
  - What is the most language in these sites?
  - What is the most news type in this dataset?
  - Show top 10 Authors with most news articles
  - Counts news data that have images or not

## • Data Description

- Columns:
  - Author. (Text)
  - Published: (Text)
  - Title: (Text)
  - Text: (Text)
  - Language: (Text)
  - Site\_url (Text)
  - Main\_img\_url (Text)
  - Type. (Text)
  - Label: (Text)
  - Title\_without\_stopwords (Text)
  - Text\_without\_stopwords (Text)
  - hasImage (Text)
- o Rows:
  - This dataset has 2043 rows.
- O Data size (no. of rows and no of columns):
  - 2043 rows and 7 columns
- Tools
  - Programs: Jupyter Notebook , Spyder, GoogleColab, MS Word, MS PowerPoint, GitHub
  - o Libraries: Pandas, NumPy, Matplotlib, Seaborn, Sklearn.
- MVP Goal
  - ➤ Applying classification models on our dataset
  - ➤ Comparing Models and choose model that has better accuracy
  - > Answering the questions
  - > Visualizing the result after applying the models.