# Information visualisation: project dataset selection

## Below you can find the 3 datasets that were selected by our group in order of preference.

## Choice 1: Global Wind Power dataset

|  |  |
| --- | --- |
| Source | <https://globalenergymonitor.org/projects/global-wind-power-tracker/> |
| Key user | Regular person with some scientific background interested in the evolution of wind power. They know high school physics likethe difference between power & energy and understand prefixes such as kilo, mega, giga, tera, ... but they are no energy expert. |
| Initial ideas | 1. GIS visuasliation of global wind energy distribution. 2. Evolution over time of installed wind power (animation) 3. What is the largest relative / absolute growth area? 4. Where are projects being cancelled? |
| Size | 26.000 observations |

## Choice 2: Billionaires dataset

|  |  |
| --- | --- |
| Source | <https://www.kaggle.com/datasets/endofnight17j03/billionaires-statistics-dataset> |
| Key user | A non-financial regular person interested in wealth distribution. They know what currency is and can distinguish millions from billions. |
| Initial ideas | 1. Graphs of wealth distribution by category 2. Geographical chart of wealth distribution 3. Net worth by industry 4. Distribution of gender, age, self-made or not, etc. |
| Size | 2540 observations in 35 features |

## Choice 3: Movie dataset

|  |  |
| --- | --- |
| Source | <https://www.kaggle.com/datasets/ashishjangra27/imdb-movies-dataset> |
| Key user | A movie enthusiast. Knows what a director or producer is. Has prior knowledge about movie ratings (PG, R, ...). |
| Initial ideas | 1. Identify trends in movie release dates and analyze their impact on revenue. 2. Analyze the relationship between budget, revenue, and popularity to determine factors that contribute to a movie's success. 3. Explore the impact of movie genres on popularity and revenue. 4. Investigate the correlation between runtime and audience engagement. 5. Visualize movie popularity over time and identify popular genres in different periods. |
| Size | 2.500.000 observations in 14 features |

## Team members

|  |  |
| --- | --- |
| Jorrit Vander Mynsbrugge | jorrit.jules.vander.mynsbrugge@vub.be |
| Mishkat Haider Chowdhury | mishkat.haider.chowdhury@vub.be |
| Ruth Vandeputte | ruth.vandeputte@vub.be |