**Data science project**

Submitted by:  
Matan Roginsky – 206328346  
Ron Hugi – 313470189  
**Research question**:

By using his tweeter account for over 11 years with a follwers base of over 88 million former president of the USA Donald J Trump influenced the American people and the rest of the world in a variety of issues. Our research question is - on which subject did Trump have the most impact regarding public opinion?

**We intend:**

To dissect all 50K+ of Trump's tweets and find out by using unsupervised learning, which cluster of tweets had the most impact on public opinion over the years. We will achieve that by counting the number of likes and shares of each cluster.

**Features:**

7 columns are mentioned below:

ID – TYPE INT

TEXT – TYPE STRING

DATE – TYPE DATE

RETWEETS – TYPE INT

LIKES – TYPE INT

IS DELETED FROM TWEETER – TYPE BOOL

IS RETWEETED – TYPE BOOL

Instances:

over 50,000 rows, each row will contain past tweets from Donald J Trump, the amount of likes they received, retweets of them, and more.

**Data sources:**

The Trump Archive – A website which archived all of Donald Trump's tweets – 56,571 tweets from 2009 up until 2021 when Tweeter blocked him from the platform.

<https://www.thetrumparchive.com/>

**Data mining methods:**

Crawling from a website: thetrumparchive.com

Libraries – BeautifulSoup4, Requests, CSV.

**Planned visualizations:**

Two dimensional visualizations including – tables, scatter plots, pie chart etc.

**Planned models**:

Unsupervised learning - clustering, elbow, silhouette.

**Validation methods**:

R^2.