**Explore Data:**

* Inspect data and it features
* Compute descriptive statistics
* Use data visualization to help identify significant patterns and trends in the data

**Questions to answer (& hypotheses to check):**

1. How does the title of an article affect its shares?
2. When are most articles published?
3. Which channel content are users more interested in reading?
4. Should the content be positive, negative or neutral?
5. Are linear models appropriate for usage here?
6. What are the most important parameters that affect shares number?

**Not done**

**Done**

**Taken (In process)**

**Tasks:**

1. Nan check + Infinity check
2. Check histograms distribution
3. Root transform to skewed cols (to fix skewness)
4. Find and plot columns with outliers (IQR analysis)
5. Do IQR filtering (?? May be not necessary 🡪 understand the reason for bimodal distribution for many features. Answer – do we need to **filter out the outliers?**)
6. Top down / down top model selection
7. Trying log transform for Y and write down the effect