**Pandas:**

**Groupby Function:** <https://www.freecodecamp.org/news/pandas-dataframe-groupby-method/>

**Remaining Topics to cover:**

Time series Analysis

**General Notes:**

* If you get a normal distribution by applying a log function to a dataset then dataset is log normally distributed.
* Cosine similarity is used for detecting the document similarity
* In case of modified Z-score using median formula

**MAD =**  median( |x-median(x)| )

**Modified Z-score =** (0.6745 \* (x-median(x)) / MAD)

**Machine Learning Notes**

* Gini Index ranges from 0 to 0.5 whereas Entropy ranges from 0 to 1.
* In huge dataset it is recommended to huge gini index due to time complexity.