Questions from Data Diggers

1. Approach: （topic choose）

ii) Find out which living condition affect happiness or women living decision:

We have the rankings (happiness and women livable cities, or based on existing articles) and possible living condition🡪we calculate the weight of each ranking and compare (Linear Regression model)

-Should we group similar factors into 1, and later analyse the correlation between them?

1. We create our own ranking of livable cities

-we collect factors (source from article: Factors\_variance.pdf)

-we collect weights (source from article: Chinese document)

-Regression model (sum of products)🡪 we have dataset and weights, we calculate the ranks.

1. If we merge 8 datasets into 1, the final set will be considered as 1 dataset or 8?
2. Should we use an ETL-like approach or manual cleansing?
3. Should we analyse the relationship between factors, after we calculate the factors weight?
4. Datasets check (If we cannot find dataset for a particular factor, can we eliminate it?)—once the datasets are fully cleaned, we will resolve the inconsistency problem, and decide our database structure.
5. Tools check – Mongo DB, Python (for processing), Scrapy, Tableau (for reporting), Github