Imagine It is a beautiful day in the city, the sun was shining bright and there was a gentle breeze in the air. You are out and about, enjoying the weather and going about their daily lives. Do you feel leisure and happy in this case? D o you want to take a photo and write some words to record your joy? Do you want to share it with your friend? [next slides]

If it is the case, you are not alone. Denissen et al. (2008) has demonstrate good weather can have a positive impact on people’s mood.

To be more specific, temperature is supposed to have a somehow big positive effect on emotion. And temperature/ wind power/sunlight have a negative impact.

Due to the limit of sample size and sample location (the authors use questionnaire to get the data about people’s mood), the authors are not able to do some further analysis (like the author cannot analyze the weather conditions across climate area since those questionnaires are only distributed in Germany!).

But with the help of Data Science! Maybe we can do more!

Our presentation will be as follows: the first part will be an overview of what we did. The second part is about how we get and process the data. The third part is training and evaluating the model. And the final part is some reflection on this project. [next side]

The first part is overview. In this project, we are try to set up a quantified model of the weather and people’s emotion.

We need two types of data in this project: 1. The first one is the weather data 2. The second one is the people’s emotion data (in this case we use posts on Weibo)

We expected that we can produce an model which can describe and predict the general public’s mood based on the weather data. [next slide]

The second part is getting and processing data. First, we look at how we get the weather data. We download the history weather data for Shanghai From 2022-03-05 to 2023-03-05 From visual crossing which is an API to download history weather data. By typing the city name (shanghai in this case),we can get the weather data including max/min/average temperature; humidity; precipitation ect. And that is what we want for this project.[next slide]

Second, we need to get the emotion data. In this project, we use posts on Weibo. Weibo is a popular Chinese microblogging platform similar to Twitter. A lot of people, news agency and government officials will post their blogs on Weibo. We believe it will be a good resource to analyze the emotion of general public.

We implement spider to grab the relevant post on Weibo. The key word is set to be “上海 天气”(Shanghai Weather) and the time range is restricted to be 2022-03-05 to 2023-03-05 which is identical to that of the weather data. We saved the text and the time of the posts to an csv file for later analysis. [next slide]

The model cannot analyze Chinses directly, what can we do? Natural language processing of course! We use stanza library for NLP. In this NLP library, the sentences are interpreted into the different integer based on the emotion they expressed. To be more specific, if the emotion the words convert is recognized to be positive, number 2 will be generated to this sentence. Similarly, 1 for neural and 0 for negative.

What if there are more than one sentences in a post which is supposed to be a very common case? Average it! [next slide]

Here is a summary of what we did: we convert a lot of posts into a list of number (I call emotion index here) and also we have its corresponding posting time as well.

We need to combine both weather data and the post data altogether to make a completed dataset.

So we just use the posting time as a key to look up the weather data sheet and add the corresponding weather data to the columns after the post data. In this case, weather data and post data can be read together.

Next, I will let my partner Rongfan to introduce the detail of our data and how to train and evaluate our model.