After graduation, I want to be a data analyst working in the Internet field. Perhaps, I may be an intermediate data analyst who is responsible for extracting correct business and product data, making reports and insightful analysis, which is important for decision making.

The dataset that I found is from kaggle and it is a dataset of predicting the click rate about display ads. And it is displayed on the website of Taobao, which is the most famous shopping apps in China. The dataset is offered by the company of Alibaba. In the dataset, they randomly sampled 1140000 users from the website of Taobao for 8 days of ad display. I will predict the probability of clicking on an ad when impressed based on the user's history shopping behavior by using this data and which user group is.

A company wants to know the CTR ( Click Through Rate ) in order to identify whether spending their money on digital advertising is worth it or not. A higher CTR represents more interest in that specific campaign, whereas a lower CTR can show that the ad may not be as relevant. High CTRs are important because they show that more people are clicking through to the website. So I plan to use this dataset to obtain the phenomenon in order to help the company to make efficient decisions. Since there are many differences in user value and consumption habits between different user groups, the advertising strategies for different user groups are also not the same. As a data analyst in the Internet field, I need to develop different delivery strategies by analyzing the relationship between different user groups and the click-through rate of ads. The preliminary question that I am going to try to answer is What is the relationship between user groups and ad click rate.

The data is obtained through Kaggle.

https://www.kaggle.com/pavansanagapati/ad-displayclick-data-on-taobaocom

Proposed timeline of work:

1) EDA & Data processing:16 Nov.

2) Modeling & Validation: 23 Nov.

3) Write up: 30 Nov