Dear all,

Your final report is due on October 20, 5:30 pm. This is a particularly important assignment (to reflect this, it is worth 10% of your grade), and I’d like to see in-depth, high-quality work. Please spend time and go deep into the presentation you are assigned in order to write a meaningful report. A superficial report would make me very, very unhappy.

Here are a few guidelines on what the report should contain.

The purpose of a report is to (i) help your employer make a decision to use the presenter’s data and (ii) tell presenters what should be changed in the presentation. You have to recommend either Update or Reject. If you’re recommending Update, (ii) it has to contain everything you would like presenters to work on before the employer accepts their data; if you’re recommending Rejection, you just have to support this assessment and try to give some constructive comments to help the presenters.

Keep in mind that the employer (and the presenters) only care about what you write to the extent that it answers (i) and (ii) above. You should write in a clear, simple fashion, and make your points with as few words as you can.

Your report should start with a brief summary of the presentation. This should not simply repeat the content in the slides but point out what you view as the key features of the data (which can be different from what the presenters emphasize). Since this is (by assumption) a ‘Update’, you should also explain why you think this data is worthy of further consideration. You can include this in the summary (say what is particularly nice or interesting about the data), or do it separately. There is typically no reason to make the summary longer than half a page or so.

Then you should make a list of what you think should be changed or clarified. You should do this so that it is clear what the main problems or suggestions are, and which comments are minor. I often have a Main comments section which includes at most 2-3 points, and an Other comments part where I put everything else. Numbering your comments makes it easier to refer to them, and imposes discipline on your writing. Especially, spend time commenting on cosmetic stuff like the appearance of graphs and slide arrangements.

Each figure must have a purpose. You have to examine the clarity of the pictures and address their professional quality. You can suggest extensions or other applications that the presenters could consider, but you should make it clear if this is something they need to do in this presentation (e.g., to convince you that their findings are robust), or if it is something that they can just mention in the Conclusions as an interesting possibility. In making these types of comments, consider what is feasible (e.g., don’t simply ask someone to redo the analysis for a different country or time-period unless the you know that the necessary data exists).

It is very important that you substantiate your comments. Simply saying “I don’t like this assumption / specification” or “the presenters should do x” is meaningless.

I expect you to pay attention to the presentation in great detail and have several specific comments. If your report is shorter than a page, you should probably put in more work.

I’m happy to discuss any of the above further, if there are any questions, please ask.

Some details: font size 11, single space, no more than three pages.

Here are the final presentation schedule:

**Thursday, October 6th**

1. Pavan Kumar Swamy Padamati, Yashodhan Damle, Ziyu Fan

**Data**: uber and lyft rides, boston (<https://www.kaggle.com/datasets/brllrb/uber-and-lyft-dataset-boston-ma>)

1. Venu Vardhan Swarna, Li Zhou, Han Gao, Wejia Yu

**Data**: Forest Coverage Data, (Original source from ‘gapminder.org’, combining with multiple datasets) (<https://www.kaggle.com/datasets/raghurayirath/population-growth-vs-forest-cover> )

1. Xiran Yao, Rixin Su, Weixin Xu,  Qinghao Yang

**Data**: Crime rate in the United States (multiple sources attached in below google sheet)

(<https://docs.google.com/spreadsheets/d/1_CvnnUILxZYwsLz64_pWEteg4JV55dzD/edit?usp=sharing&ouid=110440775032291897349&rtpof=true&sd=true>)

1. Zhengxuan Wang, Luming Peng, Sicheng Yun, Guangji Tang

**Data**: Uber & Lyft Cab prices (<https://www.kaggle.com/datasets/ravi72munde/uber-lyft-cab-prices>)

1. Zhicen Wang, Yifan Zhou, Hongxian Zhang, Emmanuel Jonathan Yalla

**Data**: US Consumer Finance Complaints (Original: Consumer Complaint Database from consumer financial protection bureau) (<https://www.kaggle.com/datasets/kaggle/us-consumer-finance-complaints>)

**Tuesday, October 11th**

1. Yijian Shen, Fangyi Wang, Li Liu, Getahun Adane

**Data**: Horse Racing Data (<https://www.kaggle.com/competitions/big-data-derby-2022/overview>)

1. Andrew Mukurazita, Daiwei Xu, Jenny Wang，Gengchen Zhang

**Data**: video game sales (<https://www.kaggle.com/datasets/gregorut/videogamesales> )

1. Jiaqing Wang, Yanxin Pang, Chenyi Wei, Ruiyang He

**Data**: Movie Dataset (<https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset>)

1. Ci Song, Rucha Patil, Vikitor Spasic

**Data**: Netflix Dataset (Main: <https://www.kaggle.com/datasets/prasertk/netflix-daily-top-10-in-us> Supplementary: <https://www.kaggle.com/datasets/shivamb/netflix-shows?resource=download>

<https://www.kaggle.com/datasets/luiscorter/netflix-original-films-imdb-scores>

)

1. Yichen Ma, Zhiwei Han, Yuxuan Zhang

**Data**: telco customer churn (<https://www.kaggle.com/code/farazrahman/telco-customer-churn-logisticregression/notebook>)