Problem Set 2

Predictive Analytics for Business Strategy

Spring 2023

Due: Sunday, February 5th

McDermott

Be sure to include all group member names on submitted assignment file and display a screenshot of any relevant output.

1. Wayfair has an idea for a new version of their website and they want to know what impact it would have on profits. They take a random sample of the users of their website and randomize who gets version A (old) and who gets version B (new).
   1. Why does Wayfair do this?
   2. Is this experiment maximizing their profits? Explain why or why not.
   3. What is the benefit of having a random sample?
   4. What is the benefit of randomization?
2. A cannabis company in the state of Washington, let’s call them Rxx, wants to make sure they have reserved the optimal amount of storage for their product at different times of year, taking into account weather patterns, seasonal demand, etc. Will they need to engage in active prediction or passive prediction to do this? Explain.
3. Suppose you have data on the daily number of taxi rides in NYC from 2000-2020. You can also see the number of Uber and Lyft rides. Uber’s share has been increasing and the taxi rides have been decreasing.
   1. If you take a random sample of this data and regress number of taxi rides on the presence of Uber, will this give you a good estimate of the causal impact of Uber’s presence on taxi rides? Explain why or why not.
4. A company called Taegyo wishes to market products specifically for consumers who are expecting a baby, such as classical music recordings, etc. They claim their products cause the baby to have higher verbal skills. Do you think their claims are likely valid? Explain why or why not?
5. Given observational data for a randomly sampled selection of adults in the U.S. on Y = annual income and X = college degree (binary), answer the following questions:
   1. Following the approach from class, what are some potential confounding factors? (State once what makes all of them potential confounding factors.)
   2. For each potential confounding factor, state whether it is a confounding factor. \*You can state once for all what makes something a potential confounding factor an actual confounding factor. \*\*It is fine to do this as a table and add the separate statements requested.
6. You are consulting a political campaign as a data scientist (after the election, so an ex post analysis). Given voter-level observational data on Y = voted for the campaign’s candidate and X = targeted with an ad, answer the following questions:
   1. Following the approach from class, what are some potential confounding factors?
   2. For each potential confounding factor, state whether it is a confounding factor.
7. If you wish to make a passive prediction, explain what kind of data you would need to have and why the other would not work.