Take-home Assignments Instructions

Assignment 1

Please solve the following problem:

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice? Please explain your reasoning in detail.

Assignment 2
About the dataset

We have pulled data about all impressions served against a subset of users on a certain platform in the month of May. Appended to this impression data, we have also added information about some of the user's favorite movie genres. Problem is, for the users where we've exposed their favorite movie genres, we've only exposed it as a blinded bucket, not what the actual movie genre is.

Here are the feature descriptions in the Data.txt file:

- Tdid: unique user identifier
- Logentrytime: time impression was served (UTC)
- Logfileid: ignore
- Site: site where the impression was served
- userHourOfWeek: user's hour of week when the impression was served (00=Sunday at 12AM in the user's time zone)
- country: country where the impression was served
- region: region where the impression was served (if in US, it's the state)
- metro: metropolitan area where the impression was served (lookup can be found here: https://developers.google.com/adwords/api/docs/appendix/cities-DMAregions)
- city: city where the impression was served (most granular geo we report here)
- devicetype: the device type of the user where the ad rendered
- osfamily: operating system family of the user where the ad rendered
- os: operating system of the user where the ad rendered
- browser: browser the user was on when the ad rendered

- FavoriteMovieGenre: user's favorite movie genre, blinded. "?????" means their preference wasn't exposed.

Questions

1. For the users where we've exposed what group they're in, can you guess the actual movie genre that group of users prefers? Be prepared to defend your answer. (i.e. does BlindedGenre1 represent action movies? documentaries? comedies? Etc.)
2. For the users where we have not exposed their grouping (i.e. "?????"), can you develop a model to predict which of the five groups they belong to? (i.e. build a model to predict which of the five blinded movie genres you think each user is associated to)
3. Can you find some users that you think would be really great friends? How confident would you be introducing them to each other? (i.e. what 2 TDIDs do you think are the best candidates for being friends, if you could introduce the two users? And how confident are you in the answer?)

Deliverable

1. Any code or workbooks you used to analyze this data. Please feel free to use whatever you're comfortable with (Excel, R, Python, whatever).
2. A lookup for blindedGenre1, blindedGenre2, ..., blindedGenre5 with your best guess at the actual movie genre it represents (question #1).
3. A lookup between TDIDs and their predicted favoriteMovieGenre (question #2).
4. No more than 10 pairings of TDIDs you think would be really good friends and your confidence level in their possible friendships (question #3).
5. Be prepared to discuss your rationale for your answers during the in-person.

Comments

Here's some general advice about approaching this problem.

1. Have fun with this.
2. Models should be as simple or as complex as you believe they should be.
3. We're just as interested in your process as we are in your actual answers.
4. Feel free to ask questions if any of this doesn't make sense.
5. The data has not been cleaned. Feel free to remove data points, but if you do, document why you did.
6. Do your best and don't overthink things.