Business Goal: Identify women that are pregnant before the release of public birth records.

Motivation:

Not only pregnant women will buy a lot of new things, they are also expected to change their shopping patterns around the birth of a child. There are chances that they will start buying everything in one store instead of buying different things at different store. Thus, if Target can send specialized ads to those pregnant women, those women might end up buying everything in Target and being a loyal customer for years.

However, since birth records are usually public, it is crucial for Target to identify pregnant women before the release of public birth records. Otherwise, pregnant women will get so many advertising flyers that the one from Target could be easily ignored.

Data Prediction Proposal:

This is a supervised classification problem.

Data Prediction goal: predict whether an individual is pregnant or not.

The target variable is pregnant or not. If the value of the target variable is yes, we know this woman is pregnant. If the value is no, we know this woman is not pregnant.

The potential features could be age, educational level, employed or unemployed, married or not, recent shopping records, whether there is a change in grocery shopping pattern, purchase of pregnancy-related nutrition supplements, how many kids one has currently, the age of the last born kids, estimated salary, what websites one recently visits frequently, the type of accounts one follows on social media, the google search records.

Model selection:

We can use classification tree to predict target value and its probability.