Springboard Data Science Intermediate with Python Capstone Project Ideas

==========

1. You are an employee of the state government and you are assigned the job of making all schools in CA get 9+ out of 10 rating. You have a few data-sets and ratings and you analyze
a) What makes a school district great one?
b) Come up with a predictive model to guess which of the n rating bins school falls into. Bins can be 3 or 4 in number, like 0 to 4, 4.1 to 7, 7.1 to 10
c) Suggest measures to bring the lower rated ones gradually to 9+ rating zone

Factors that might influence

- Income levels of households,
- Crime rate (this factor is interesting, because it could be cause effect confusion)
- Population(?),
- Population density,
- Education level of adults
- Occupation types (business vs salaried), sub-types (engineer, banker etc)
- Proximity to metropolitan (like most cities within 50 miles of LA have better ones?)
- Race
- Age distribution of residents

Another related question might be the consistency in rating over years and what causes it.
We can perhaps think of ways to increase the scope. Rating data can be found in greatschools.org. Demographic data is available through census website.

2. You have just landed in USA from a foreign country and you are trying to figure out a lot of stuff. One of them is which car to buy (or should I really buy?)
Wouldn't it be wonderful if given some data about you and place that you are in, your computer tells you precisely which car to buy, or rather give you list of 3 cars from which you make final choice?

Here's the factors to consider:

a) First question to answer is, does it even make sense to buy: Uber/biking/public transportation might be an option

- How long he plans to stay (a few years or forever?)
- Average distance to be covered daily/weekly
- Availability of bike trails in the daily commute route
- Availability of public transport
- Is it a place where winter is strong and can't take out car in certain season?
- Does he enjoy driving and hence wants to own?
- How many times in a year plans a road trip
- Single/Married
- How many kids? Any infants?

b) Second question: How to select?

- What is the budget
- Does he want used or new car?

- Color/Model/Make/Year preferences - Doesn't mean search will be limited there.
- MPG requirements
- Gas/Hybrid/Fully electric?
- What is the resale value of a car after 5/7/10 years? 10 years down the line will there be any takers for gas cars?
- Is it high end/low end? Demand in the market
- Maintenance cost per year (service, insurance etc)

Not sure how to get the data to train the model and predict a match for every new person

3. Talking of cars, one more idea: One of my friends always claims that she is scared of red cars in freeways because it indicates aggressive mind and hence chances of getting into collision is more. I thought that this and its extended problem can be studied
a) Is it true that red cars get into more collisions than other cars?
b) More general question: is there a correlation between color/model/makes of the car and the collisions that they get into?

Got dataset obtained by a Chicago resident from Illinois Department of Transportation (IDOT) here - suggested by mentor:
https://github.com/stevevance/Chicago-Crash-Browser/blob/master/DATA.md

==========

History:

1. During first mentor call, the goal was set for the next week to come up with datasets/ideas to work on.
2. The next week, that is, during second mentor call, we discussed a few of UCI datasets, but didn't choose because:
a) They are too academic
b) Not challenging enough to attract attention from industries.
Decision: Should define a problem that is preferably not looked into so far, with decently big in number of samples and features.
3. Before the third call, through email, several ideas were discussed. In the third call, we were inclined towards idea 3 above. Since this is just the third week, we are still open to new ideas.

Date: 12/21/2017