Name: Chang Liu

# Summary of activities:

Repository link: <https://github.com/changliu8888/Amazon-driver-coupon>

I finished all the suggested assignment questions and did self guided investigation. What I did is the following steps:

1. Look inside the database and understand the structure. Total 56.8% coupons accepted. That is pretty strong case for the businesses.
2. Look at different filtering groups and determine their chance of buying a coupon.
   1. I created a new data frame dfbar for drivers who had visibility of Bar coupons. 15.9% coupons were the Bar type. 41% of Bar coupons were accepted on the basis of Bar coupons given out
   2. Out of the bar coupons accepted, 47% from people who to go bars 3 times or less. 9.8% for people who go to bars more often.
   3. Drivers who go to bar more than 1 per month and over 25, the acceptance case is 397, compared with 827 bar coupons accepted. This is about half.
   4. 574 coupons accepted for people who go to bars more than once a month and had passengers that were not a kid and had occupations other than farming, fishing, or forestry. This category seems to be a good focused hit, greater than 50% acceptance.
   5. My hypothesis is this. So far, the people who use the coupons the least among the three comparison groups are those who go to bars once a month and under age 30. The people most likely to say yes among 3 comparison groups are drivers who have no passengers as kids and not widowed are least likely.
3. Independent investigation
   1. I changed the income to a new number column, incomelow, in a dfnew dataframe.
   2. I created different histogram plots and count plots with hue to see which categorical sector has high acceptance.
   3. Groups with high percentage of acceptance are
      1. Carry out and take away coupons
      2. Restaurant less than 20 coupons
      3. NO urgent places type drivers
      4. Some college degree or no degree drivers.
      5. People who work in architecture industry, education/training and library, and health care support seem to have high rates of accepting coupons.
      6. Singles have higher chances of accepting coupons.
      7. People with income of around 40000 has high chance of accepting coupons.
   4. These groups seem to me the highest priorities for targeting coupon ads.



# Code, visualization techniques I explored

I discovered a few new methods not mentioned. These include count plot, and composite query. Using my own way I turned the income to a number variable.

1. The income is categorical, not numerical. I created a new column called “incomelow”. This contains numbers which is basically average of the range. Using this technique I turned income into a number category.
2. Instead of trusting a hunch and query on a complex condition, I want to find the biggest categories that has high acceptance rate and large number of drivers. I also want to find the opposite. I built a number of histograms to gain insights.
3. I looked at histogram of coupon types, occupation and determined the following groups has the best rate of buying a coupon (any coupon)
   1. By occupation
   2. By income level
   3. By establishment type
   4. etc

# Overall findings:

The following category has a high chance if my budget for coupon advertising is limited:

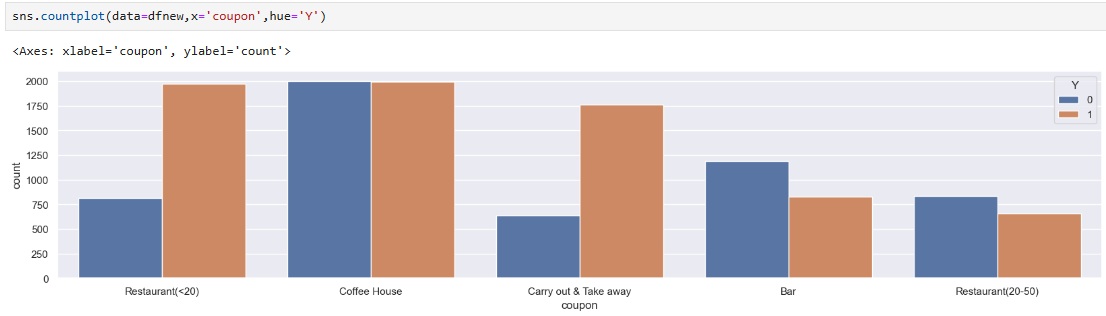
1. Singles and married with partner has >50% chance of accepting.
2. “Income equals 40000” audience has a high chance.
3. By occupation, architects, art design and entertainment have higher chance/
4. People going to no urgent place has higher response rate.
5. Low-cost restaurants and carryout has higher chance of responding. Bar and high-cost restaurants have lower chance.

The categories I would not bother heavy bombardment initially are

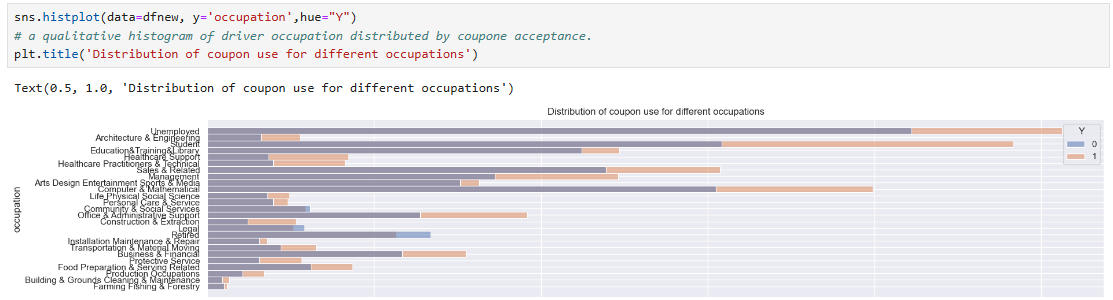
1. Widowed drivers. Their population is low.
2. Divorced drivers. Their acceptance is lower (just slightly over 50% and their number is low)
3. Retired drivers. Their chance of accepting is less than 50%, lower than average.
4. Restaurant over 30 dollars is not a good category to target.

Here are some sample graphs. For details please reference to the notebook in [this depository](https://github.com/changliu8888).

For example, a side by side comparison and histogram by different types of establishment.



Here is a distribution histogram by different occupations, along with accepted/not accepted comparisons. Placing the graph horizontally was useful to reduce clutters.



# Next steps and recommendations

The coupon acceptance rate is very high. To be honest, I am a bit suspicious about the high rate of acceptance. However, this kind of data is very important for business.

As next step, I would like to suggest

1. Connect between the customer profile and whether they actually used the coupon and made spending.
2. The ROI (return on investment) for such campaigns.