**Coupon Project**

To start with, I explain the methodology followed and later I get into my findings.

**Setup**

1. First, I loaded the data into dataframes and analyzed the data for outliers/incompleteness.
2. I noticed that ‘Car’ column did not contain data for the most part (99%+) so I dropped that column. All other columns contained data at healthier rates, so decided to keep them.
3. Next, I looked at other nulls in the data. I decided to clean the data further and removed the rows containing any NaN. From the original 12684 rows, I could retain 12079 rows (over 95%) of the data, even after deleting the rows containing NaNs.
4. Now that I have a clean data set without Nulls, I wanted to fix the columns containing datatypes that were not correctly defined.
5. . However, I saw some columns that had strings instead of numbers e.g. ‘age’. I converted this to numeric. Less than 21 and above 50, I made them as 20 and 50 in the dataframe.

**Findings**

1. 56.9% accepted the coupons when offered
2. ‘Carry out & Takeway’ and ‘Restaurants<20’ has the highest coupon acceptances. This was expected as coupon clippers are usually price sensitive (hypothesis). Typically carry outs and restaurants less than $20 are on the affordable side and hence appeal to people who use coupons. These two categories had acceptances above 70% while more expensive restaurants ($20 to $50) had 45% acceptance rates approximately.
3. Another hypothesis I had was, if the weather is nicer, there would be more people accepting coupons and make use of it. Based on my analysis, coupons acceptances were highest when temperature was above 80 degrees. However, there wasn’t much difference in acceptance rates when temperature was 30 or 55 degrees.
4. About 41% of Bar coupons were accepted
5. When bar coupons were presented to people who visit bars, 3 or more times, acceptances were quite high at 76%, against 37% for people who visited bars less than 3 times.
6. Probability of accepting coupon if age is above 25 and they visit bars more than once a month is 69%, as against 34% for others.
7. Came up with the following hypothesis:
   1. Null hypothesis: Accepting coupon when users go to bar more than once and their age is over 25 is same as all other drivers.
   2. Alternate hypothesis: people who visit a bar more than once and are aged over 25 is more likely to accept coupon
8. Tested the null hypothesis and alternate hypothesis using scipy.stats and p-val. In my case null hypothesis was accepted due to an issue p-value generation. I received a p-value of NaN due some issue. If a correct p-value was obtained, alternate hypothesis might have been correct.
9. People who are below 30 and go to bars more than once a month, had an acceptance of 72%
10. Did independent analysis with ‘carry out’ data. Found 80% of coupons were accepted by people who are in ‘Management’ occupation. Closely followed by Sales and Computer professionals.
11. Assumed unemployed folks would save the money and eat at home more. However, over 75% of the coupons issues to this group used the coupon.

**Next Steps**

1. We could create correlation matrices to find variables that are strongly correlated and whether the correlation is positive or negative
2. We could analyze all varieties of coupons restaurants, bars etc. and have different promotions for different age group, income level, occupation type etc.
3. We could use the sample data provided to create a model to determine who would be more likely to accept coupons if presented.