Context

Imagine driving through town and a coupon is delivered to your cell phone for a restaurant near where you are driving. Would you accept that coupon and take a short detour to the restaurant? Would you accept the coupon but use it on a subsequent trip? Would you ignore the coupon entirely? What if the coupon was for a bar instead of a restaurant? What about a coffee house? Would you accept a bar coupon with a minor passenger in the car? What about if it was just you and your partner in the car? Would weather impact the rate of acceptance? What about the time of day?

Obviously, proximity to the business is a factor on whether the coupon is delivered to the driver or not, but what are the factors that determine whether a driver accepts the coupon once it is delivered to them? How would you determine whether a driver is likely to accept a coupon?

Overview

The goal of this project is to use what you know about visualizations and probability distributions to distinguish between customers who accepted a driving coupon versus those that did not.

Data

This data comes to us from the UCI Machine Learning repository and was collected via a survey on Amazon Mechanical Turk. The survey describes different driving scenarios including the destination, current time, weather, passenger, etc., and then ask the person whether he will accept the coupon if he is the driver. Answers that the user will drive there ‘right away’ or ‘later before the coupon expires’ are labeled as ‘Y = 1’ and answers ‘no, I do not want the coupon’ are labeled as ‘Y = 0’. There are five different types of coupons -- less expensive restaurants (under 20), coffee houses, carry out & take away, bar, and more expensive restaurants (

20 - $50).

Data Description

Keep in mind that these values mentioned below are average values.

The attributes of this data set include:

User attributes

Gender: male, female

Age: below 21, 21 to 25, 26 to 30, etc.

Marital Status: single, married partner, unmarried partner, or widowed

Number of children: 0, 1, or more than 1

Education: high school, bachelors degree, associates degree, or graduate degree

Occupation: architecture & engineering, business & financial, etc.

Annual income: less than $12500, $12500 - $24999, $25000 - $37499, etc.

Number of times that he/she goes to a bar: 0, less than 1, 1 to 3, 4 to 8 or greater than 8

Number of times that he/she buys takeaway food: 0, less than 1, 1 to 3, 4 to 8 or greater than 8

Number of times that he/she goes to a coffee house: 0, less than 1, 1 to 3, 4 to 8 or greater than 8

Number of times that he/she eats at a restaurant with average expense less than $20 per person: 0, less than 1, 1 to 3, 4 to 8 or greater than 8

Number of times that he/she goes to a bar: 0, less than 1, 1 to 3, 4 to 8 or greater than 8

Contextual attributes

Driving destination: home, work, or no urgent destination

Location of user, coupon and destination: we provide a map to show the geographical location of the user, destination, and the venue, and we mark the distance between each two places with time of driving. The user can see whether the venue is in the same direction as the destination.

Weather: sunny, rainy, or snowy

Temperature: 30F, 55F, or 80F

Time: 10AM, 2PM, or 6PM

Passenger: alone, partner, kid(s), or friend(s)

Coupon attributes

time before it expires: 2 hours or one day

Data Preparation / Cleaning

The initial step in cleaning the dataset involves removing duplicate entries.

Code: data.drop\_duplicates(inplace=True)

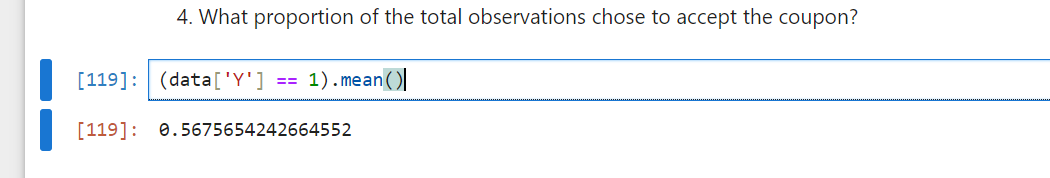
Further data cleaning, such as handling null values and converting data types, will be conducted during the analysis phase when specific columns are selected.

Summary of Findings & Questions to be answered

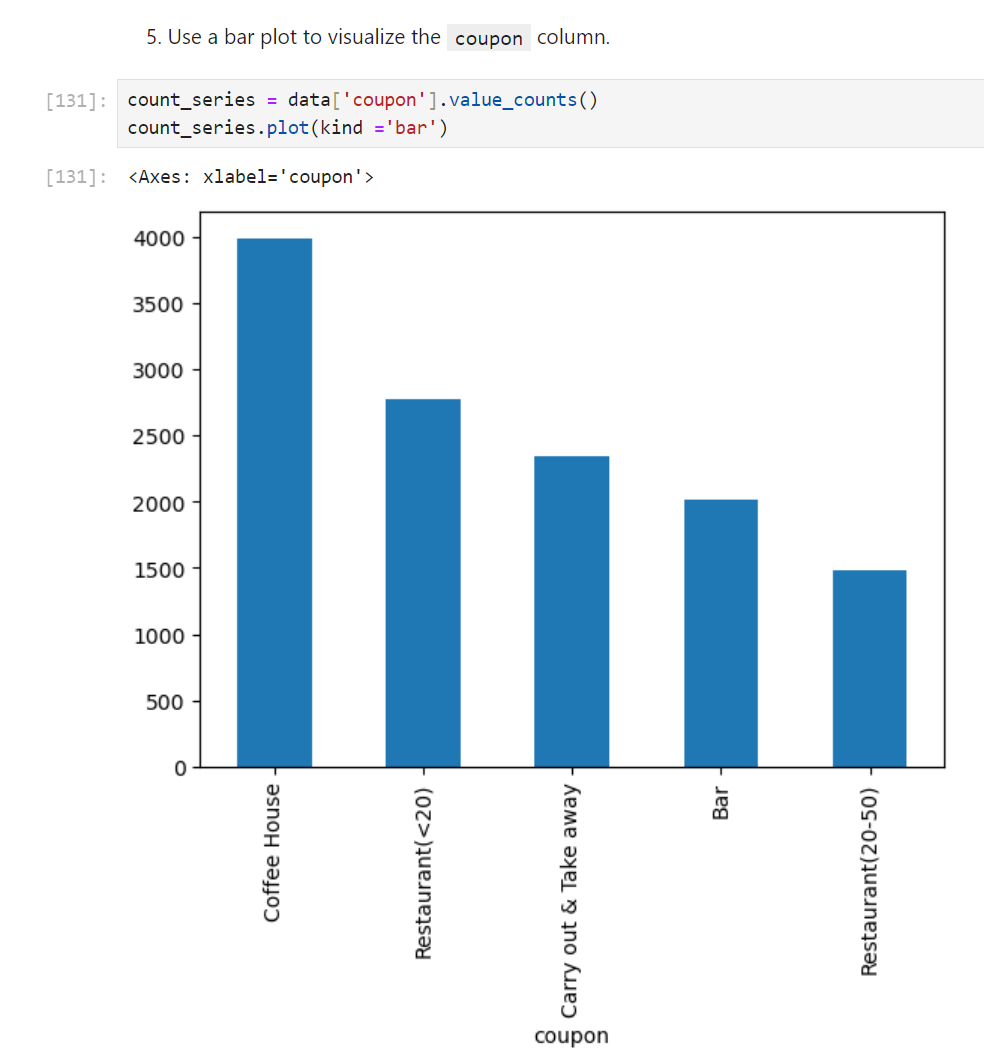
**4. What proportion of the total observations chose to accept the coupon?**

Code: (data['Y'] == 1).mean()

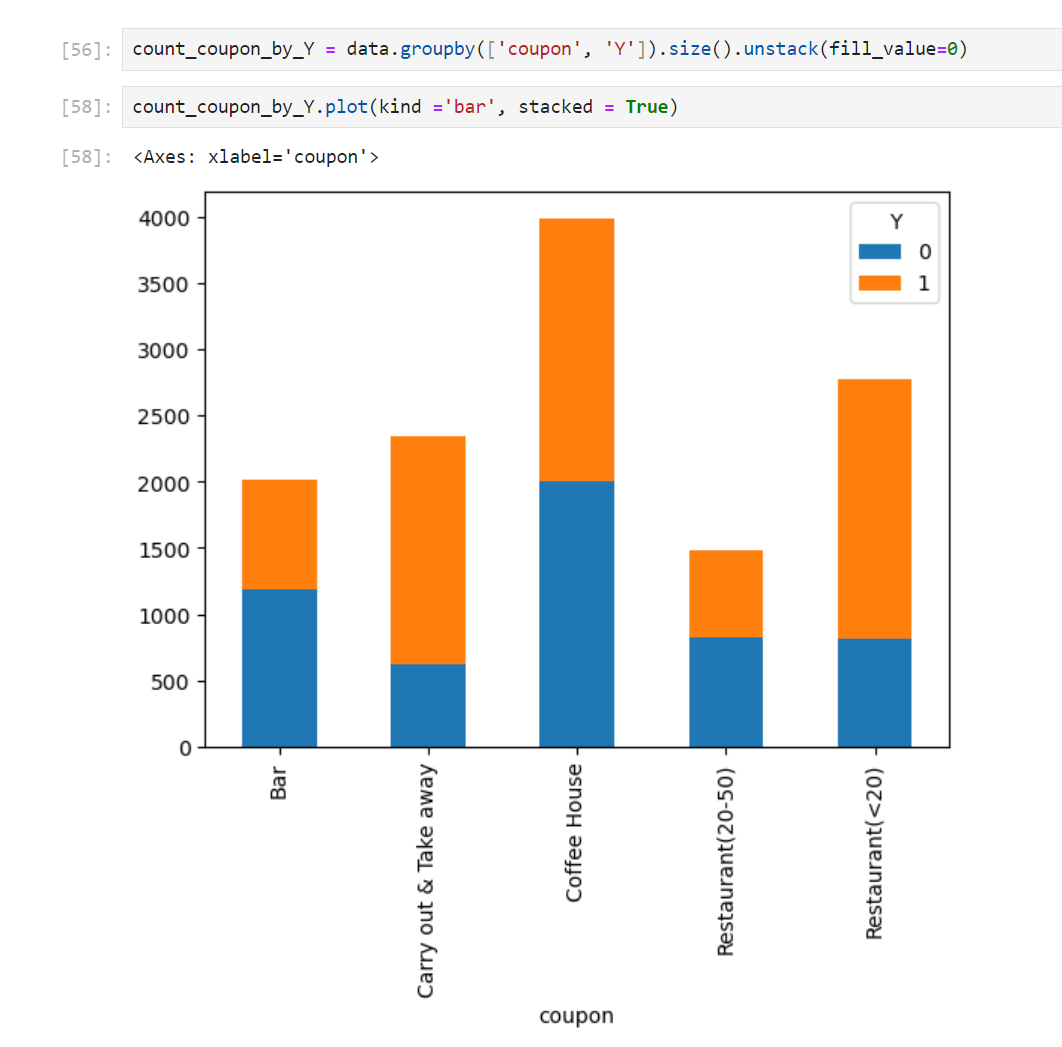
Answer: 57.675%



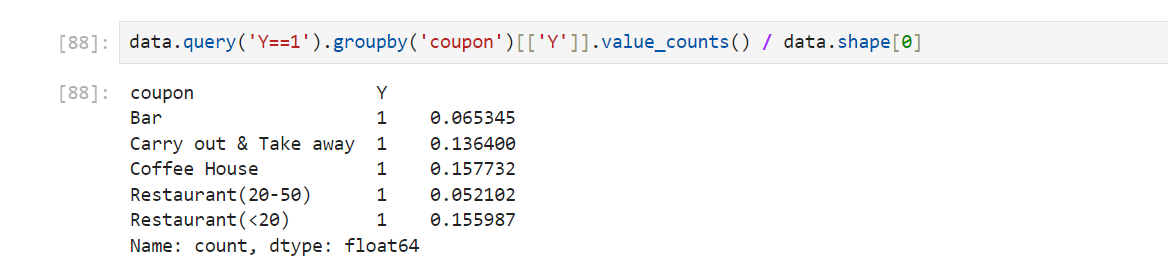
**5. Use a bar plot to visualize the coupon column.**



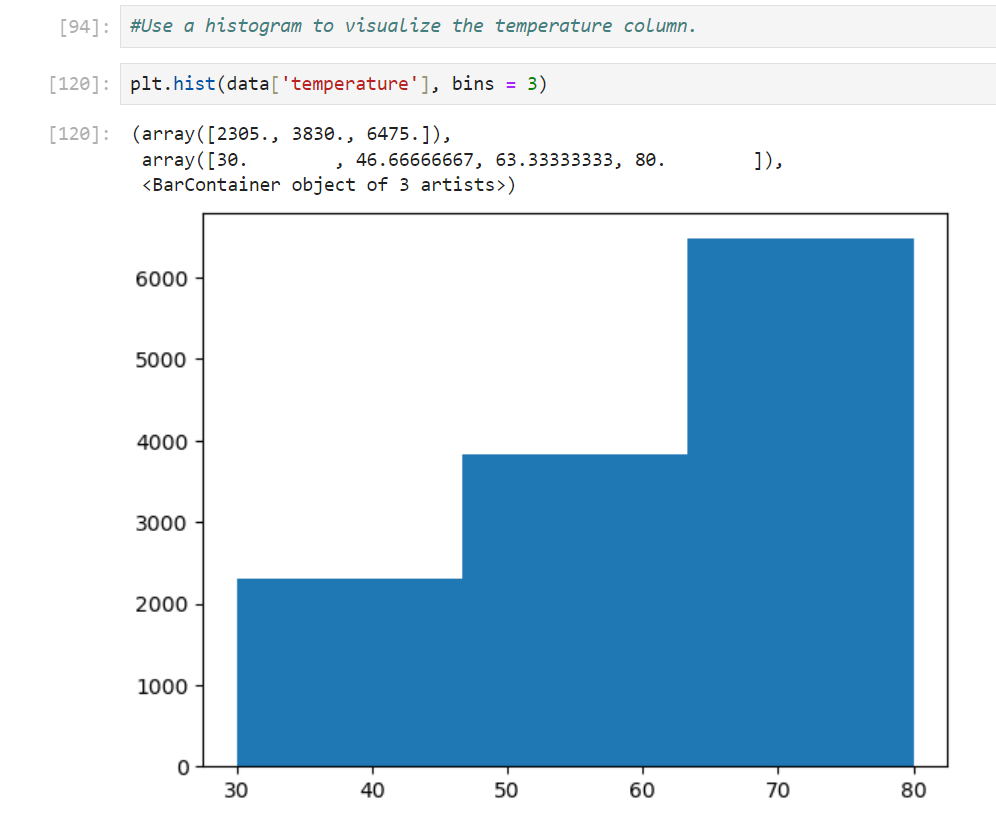
In addition to bar plot, we can use a stacked bar plot to include Coupon Acceptance (1 = Yes, 0 = No) to visualize the proportion of acceptance in each category



The acceptance rate of each coupon category is calculated below



**6. Use a histogram to visualize the temperature column.**



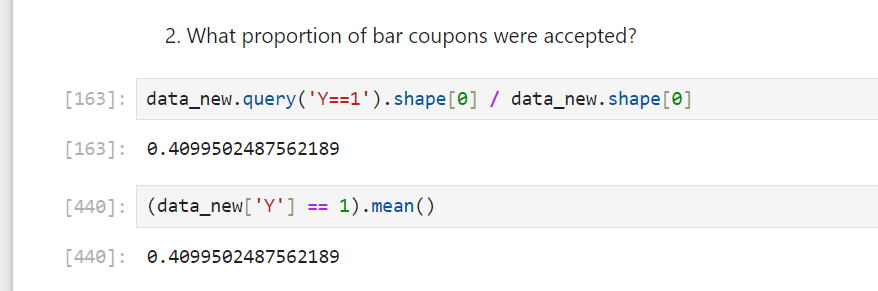
**Investigating the Bar Coupons**

In this section of analysis, we will be only looking at Coupon = Bar coupons

**2. What proportion of bar coupons were accepted?**

40.995 % of bar coupons were accepted.

In the code below, I applied two different calculations to achieve the same result.

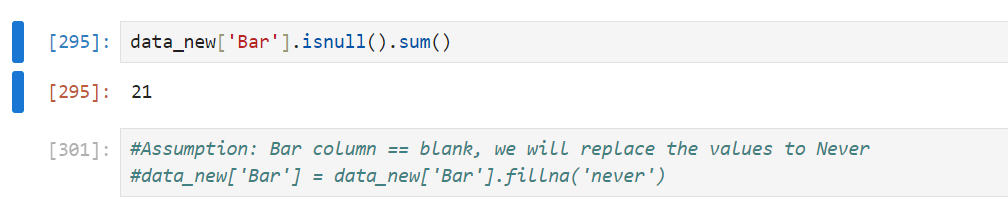


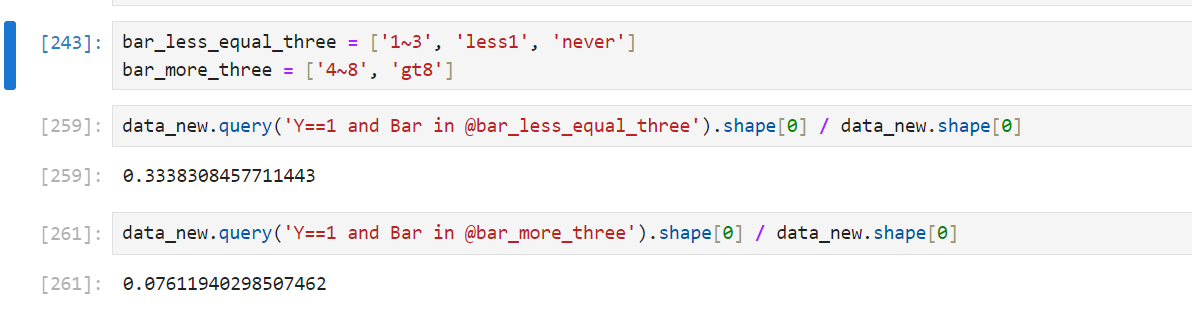
**3. Compare the acceptance rate between those who went to a bar 3 or fewer times a month to those who went more.**

Since this will be the first time using the Bar column, we need to first to analyze this column.

This column contains 21 null cells. In this case, we will assume that null values reply to driver NEVER been to bar in month.

Therefore, we will replace Null values to Never





The acceptance rate for those who went to a bar 3 or fewer time is 33.38 %, which is significantly higher than those who went more than 3 times in month, 7.61%

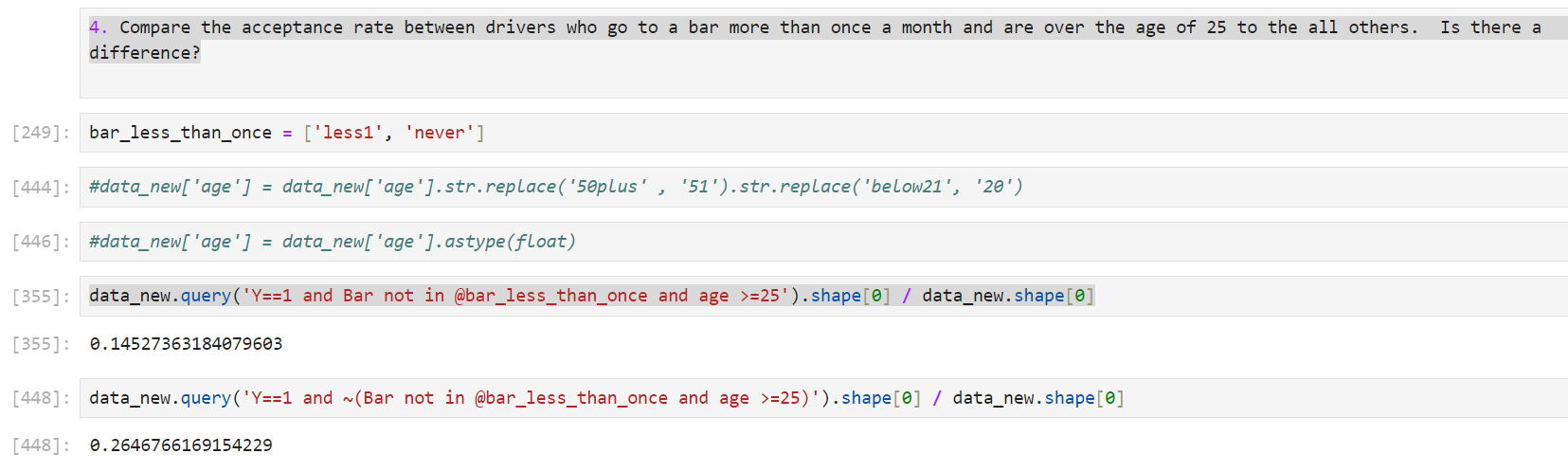
**4. Compare the acceptance rate between drivers who go to a bar more than once a month and are over the age of 25 to the all others. Is there a difference?**

In this analysis, we will need to use the Age column. In order to manipulate ages as numbers, we will need to first convert age group into Number.

In this conversion, I convert the age of below21 to 20, and 50above to 51.

Once all values are in numeric form, we can convert the data type to number from text.

The acceptance rate between drivers who go to a bar more than once a month and are over the age of 25 is 14.52 %, the acceptance rate of the rest would be 26.47%

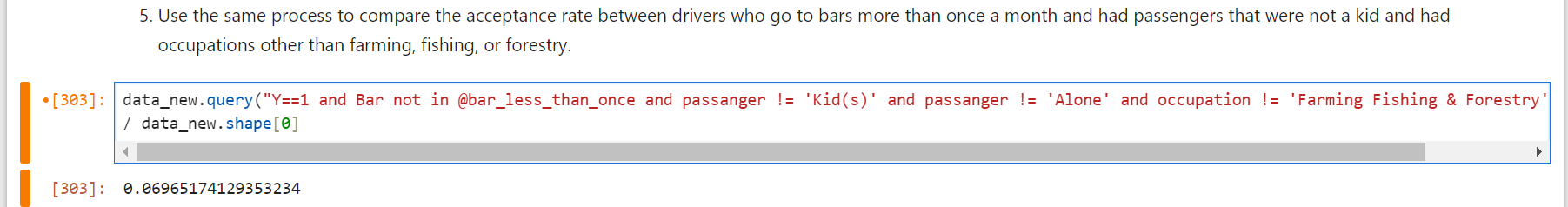


**5. Use the same process to compare the acceptance rate between drivers who go to bars more than once a month and had passengers that were not a kid and had occupations other than farming, fishing, or forestry.**

In this analysis, since the “had passengers that were not a kid” is not very clear. We make an assumption that Driver must had passengers, and can not be kid.

Therefore, we will exclude drive who has Kids and Alone driver

The acceptance rate is 6.96 %



**6. Compare the acceptance rates between those drivers who:**

* **go to bars more than once a month, had passengers that were not a kid, and were not widowed \*OR\* 6.96%**
* **go to bars more than once a month and are under the age of 30 \*OR\* 12.39%**
* **go to cheap restaurants more than 4 times a month and income is less than 50K. 7.76%**

The results are calculated correspondingly,



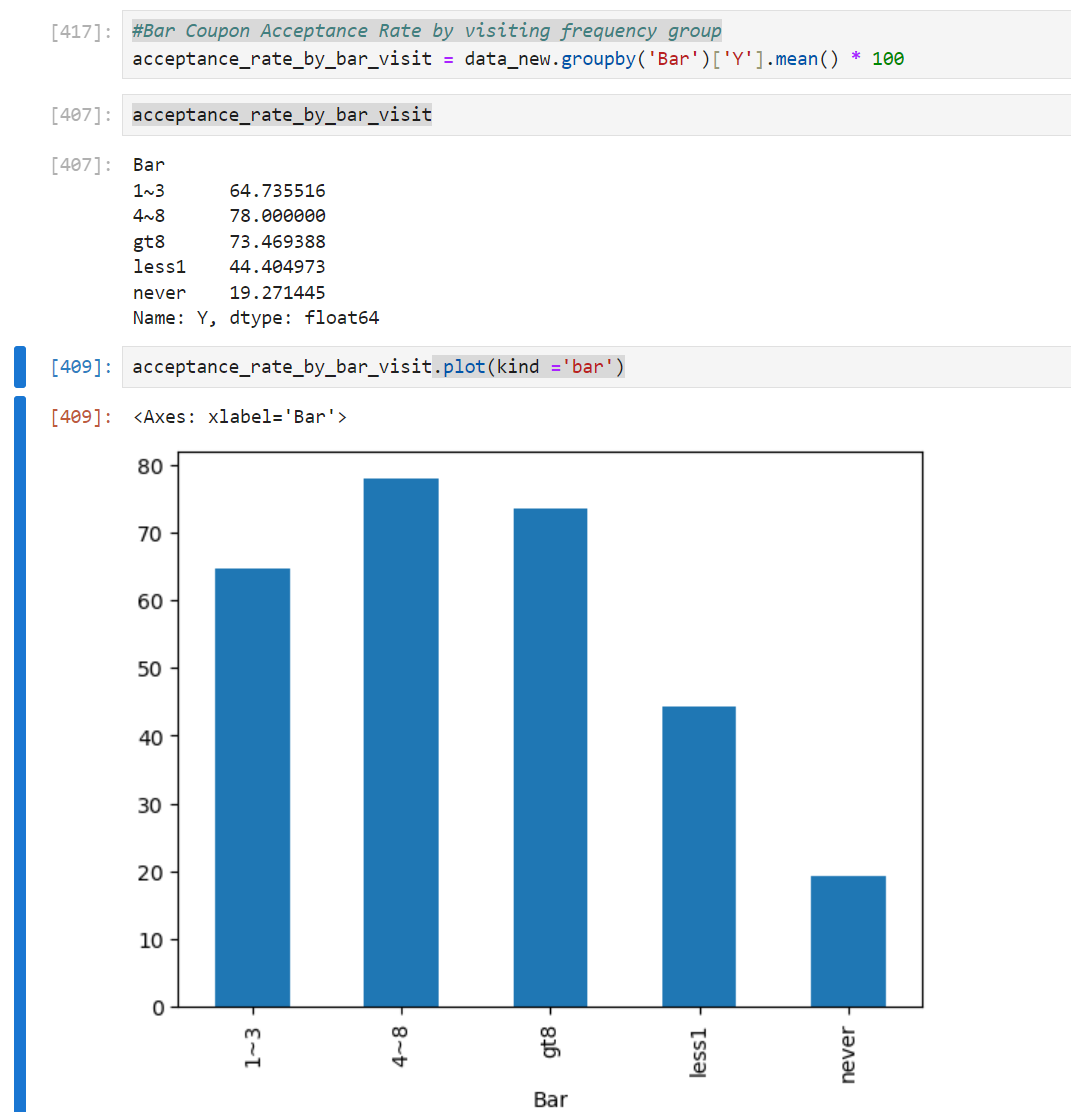
**7. Based on these observations, what do you hypothesize about drivers who accepted the bar coupons?**

Based on the observation from the above calculation, we can determine that driver who visited the bar more often tended to have higher coupon acceptance rates

**Independent Investigation**

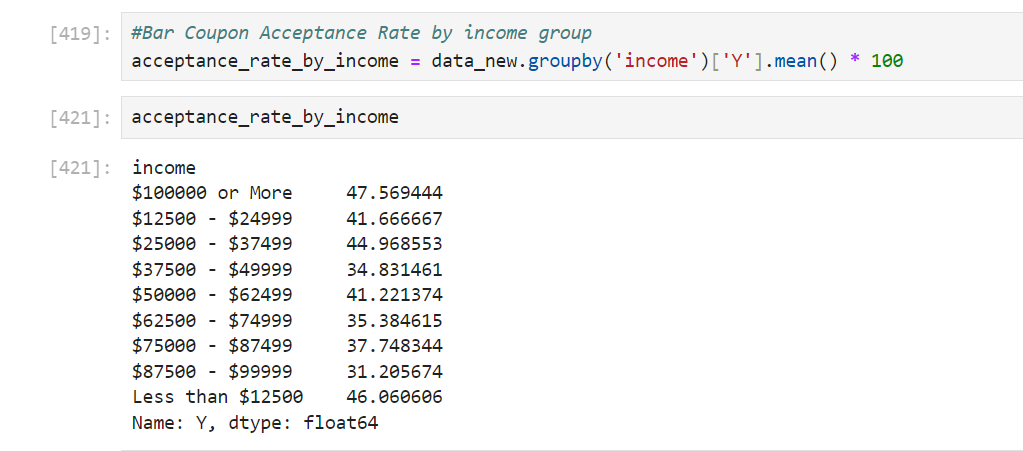
In this independent investigation exercise, I would like to look at

#Bar Coupon Acceptance Rate by visiting frequency group



As the result, drivers with high frequency of bar visiting tended to have higher coupon acceptance rate.

Another analysis I thought we be interesting to investigate is acceptance rate vs income



However, the acceptance rate of each income group is very closed. We are not able to see any significant impacts of income vs rate.