Ken Motoyama (ken.motoyama@gmail.com)

Assignment 5.1: Will the Customer Accept the Coupon

Exploring Bar Coupon Acceptance

Data Exploration

Decide what to do about your missing data

Examining the coupon data set, I noticed missing data in a column called ‘cars’ as it only had about 108 rows vs 12K+

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12684 entries, 0 to 12683
Data columns (total 26 columns):
 #   Column                Non-Null Count  Dtype 
---  ------                --------------  ----- 
 0   destination           12684 non-null  object
 1   passanger             12684 non-null  object
 2   weather               12684 non-null  object
 3   temperature           12684 non-null  int64 
 4   time                  12684 non-null  object
 5   coupon                12684 non-null  object
 6   expiration            12684 non-null  object
 7   gender                12684 non-null  object
 8   age                   12684 non-null  object
 9   maritalStatus         12684 non-null  object
 10  has_children          12684 non-null  int64 
 11  education             12684 non-null  object
 12  occupation            12684 non-null  object
 13  income                12684 non-null  object
 14  car                   108 non-null    object
 15  Bar                   12577 non-null  object
 16  CoffeeHouse           12467 non-null  object
 17  CarryAway             12533 non-null  object
 18  RestaurantLessThan20  12554 non-null  object
 19  Restaurant20To50      12495 non-null  object
 20  toCoupon_GEQ5min      12684 non-null  int64 
 21  toCoupon_GEQ15min     12684 non-null  int64 
 22  toCoupon_GEQ25min     12684 non-null  int64 
 23  direction_same        12684 non-null  int64 
 24  direction_opp         12684 non-null  int64 
 25  Y                     12684 non-null  int64 
dtypes: int64(8), object(18)
memory usage: 2.5+ MB
```

When exploring the column more, I saw the following

```
# weird that the car column has very few rows compared
# to the rest of the data set

df['car'].value_counts()
```

```
car
Scooter and motorcycle                      22
Mazda5                                      22
do not drive                                22
crossover                                   21
Car that is too old to install Onstar :D    21
Name: count, dtype: int64
```

I decided not to use this column in my analysis

All Valid Data

I also was curious if all the data points could be used and noticed the following when exploring the data

```
# are there potential underage drinkers in the data set accepting coupons?
# Yikes, looks like it

df[df['age'] == 'below21']['Y'].mean()
```

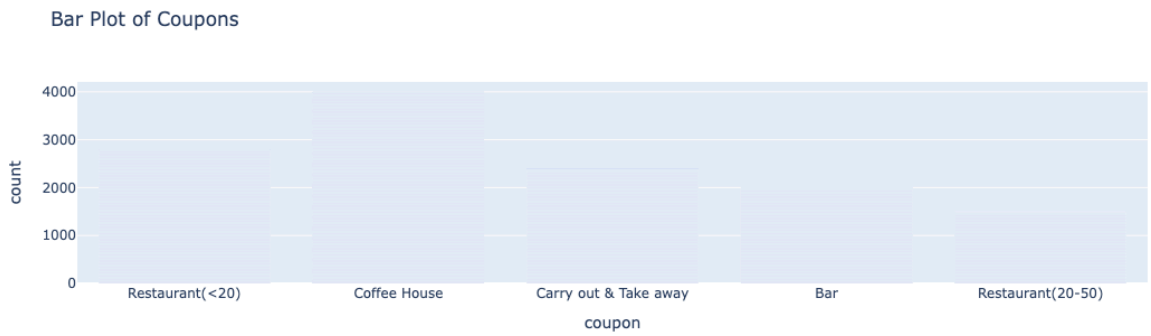
```
0.6343692870201096
```

Essentially underage people were accepting Bar coupons. Based on this, I decided to remove the ‘below21’ population from the dataset when investigating Bar coupons.

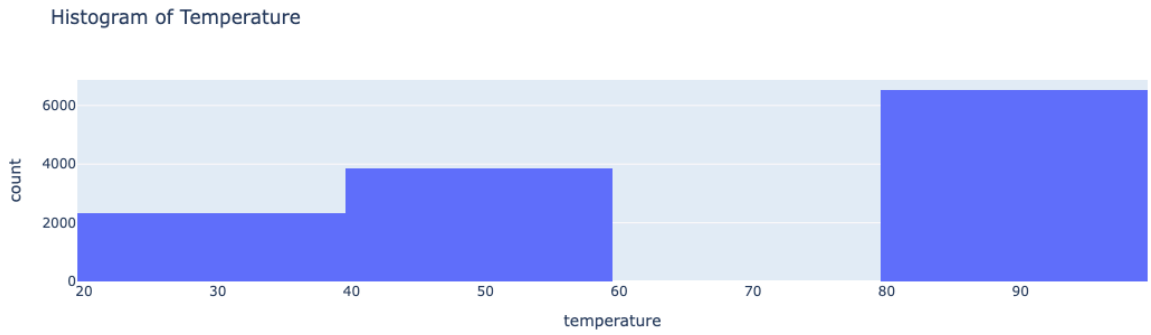
What proportion of the total observations chose to accept the coupon?

From the entire data set, the proportion of acceptance for the coupon was 0.57

Use a bar plot to visualize the coupon column.



Use a histogram to visualize the temperature column.



What proportion of bar coupons were accepted?

The proportion of bar coupons that were accepted for those of age was 0.41

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

The acceptance rate for fewer than 3 bars per month is 0.37 and for 3 or more bars per month is 0.77

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?

The acceptance rate for people who go to a bar more than once a month and are over the age of 25 is 0.695 and for everyone else is 0.330 or about 2.10 times the difference.

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.

The acceptance rate for people who go to a bar more than once a month and had a passenger that was not a kid and occupations outside of farming, fishing, or forestry is 0.715

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 go to bars more than once a month and are under the age of 30 OR go to cheap restaurants more than 4 times a month and income is less than 50K.

Drivers that go to bars more than once a month, had passengers that were not a kid, and were not widowed accepted at a rate of 0.715. Drivers that go to bars more than once a month and are under the age of 30 accepted at a rate of 0.731. Drivers that go to cheap restaurants more than 4 times a month and whose income is less than 50K accepted at a rate of 0.465.

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

Generally, I would hypothesize that drivers who accepted bar coupons tend to go to bars at least once per month.

Free Exploration of the Data

I wanted to explore a hypothesis that as income goes up, the more likely a person would be able to go to a bar, and thus the more likely they would accept a bar coupon. I ran some quick analysis on stats

Acceptance rate for all “Bar’ values

average for never is 0.156
average for less1 is 0.419
average for 1~3 is 0.649
average for 4~8 is 0.785
average for gt8 is 0.810

The number of times an individual visits a bar each month definitely has an influence on the likely acceptance rate, essentially the more times someone visits a bar each month, the higher the acceptance rate.

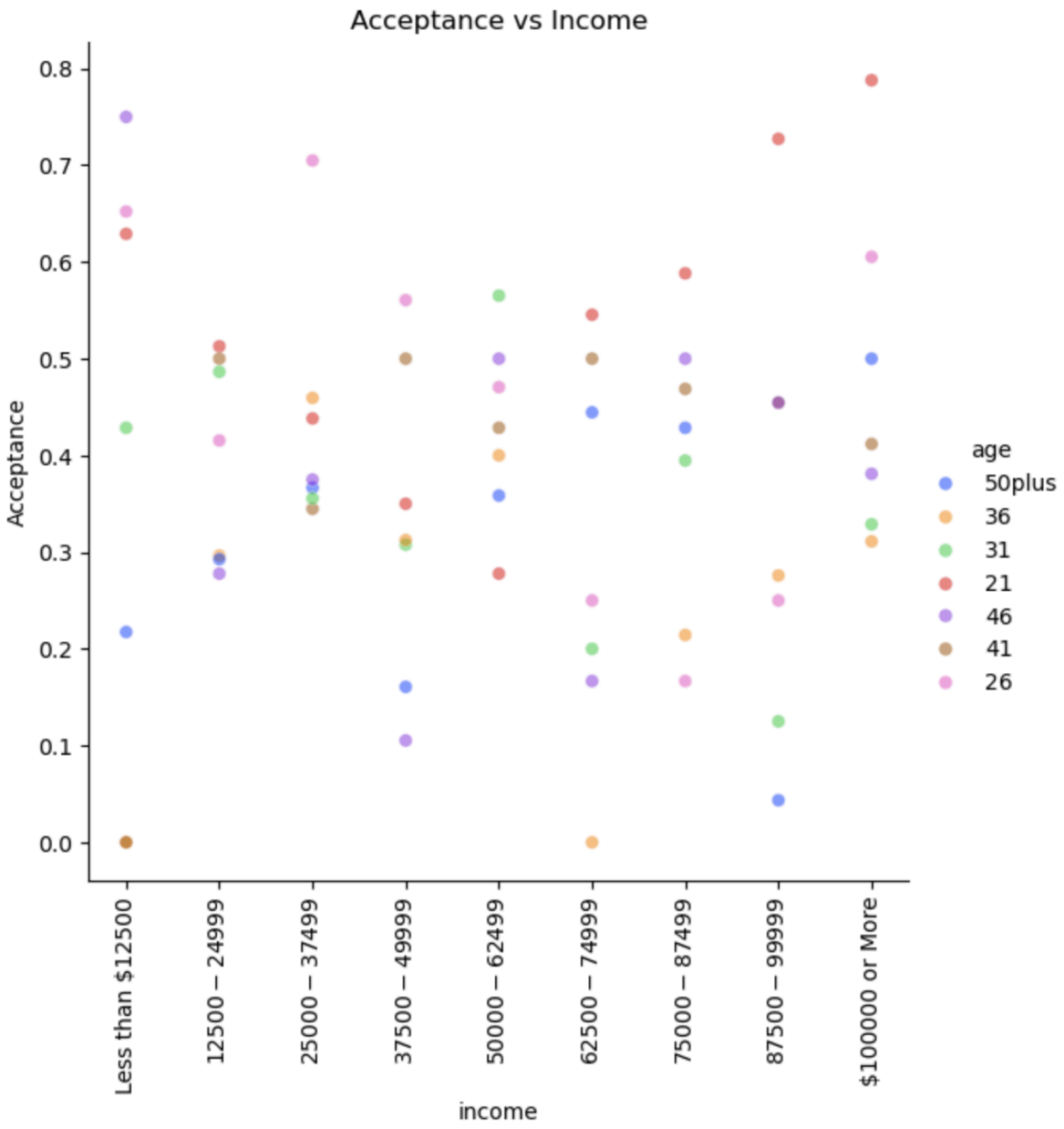
Acceptance rate for all incomes

average for Less than $12500 is 0.603
average for $12500 - $24999 is 0.519
average for $25000 - $37499 is 0.639
average for $37500 - $49999 is 0.557
average for $50000 - $62499 is 0.591
average for $62500 - $74999 is 0.431
average for $75000 - $87499 is 0.449
average for $87500 - $99999 is 0.484
average for $100000 or More is 0.618

Income doesn’t seem to be a great indicator of the likelihood someone will accept a coupon. That being said, I decided to explore more to see if maybe age influences acceptance rate.

Deep dive into Acceptance rate for all incomes

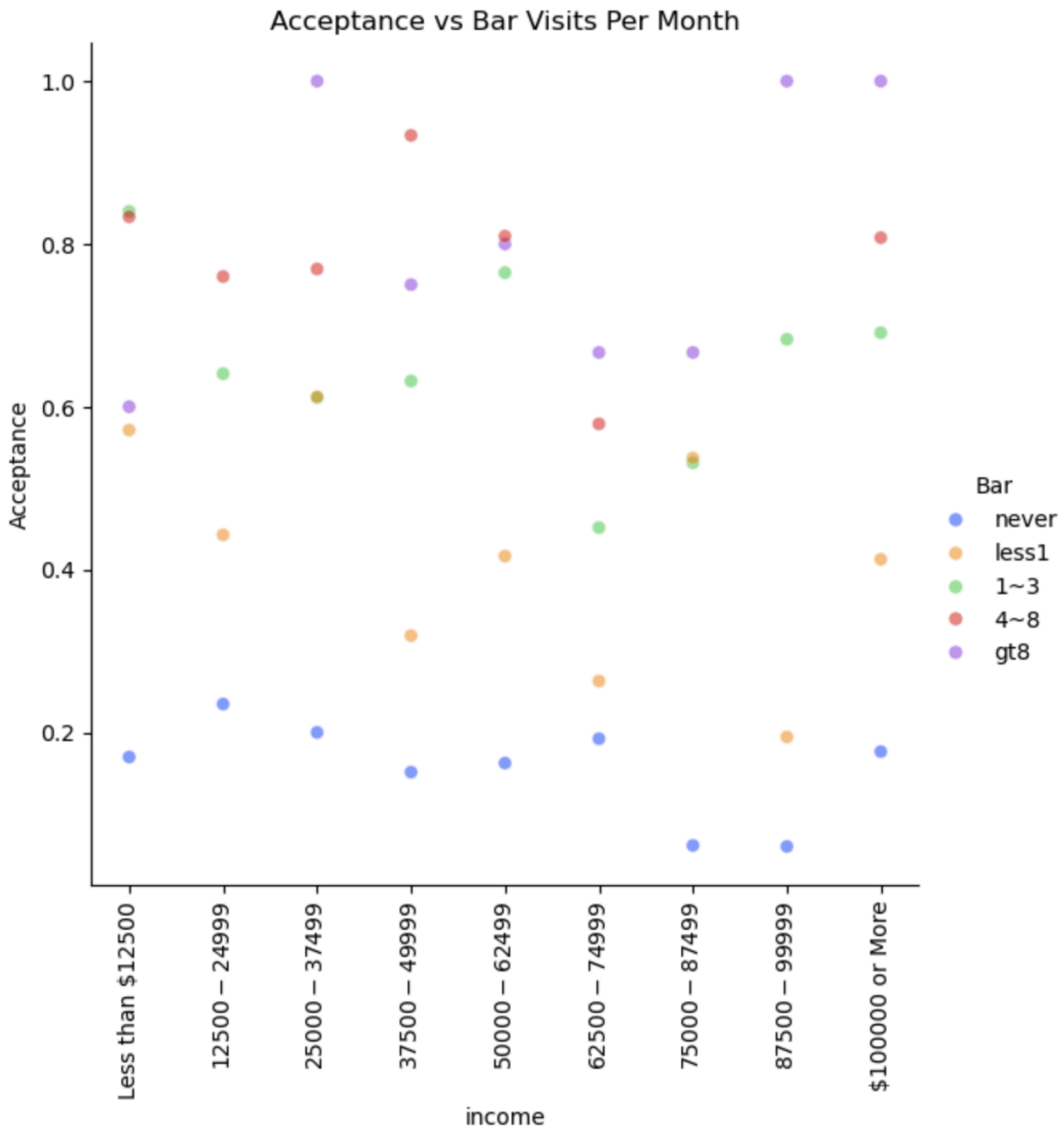
I used seaborn to add the age dimension in a scatter plot of acceptance rate vs income.



No real discernible relationship could be inferred. Some observations

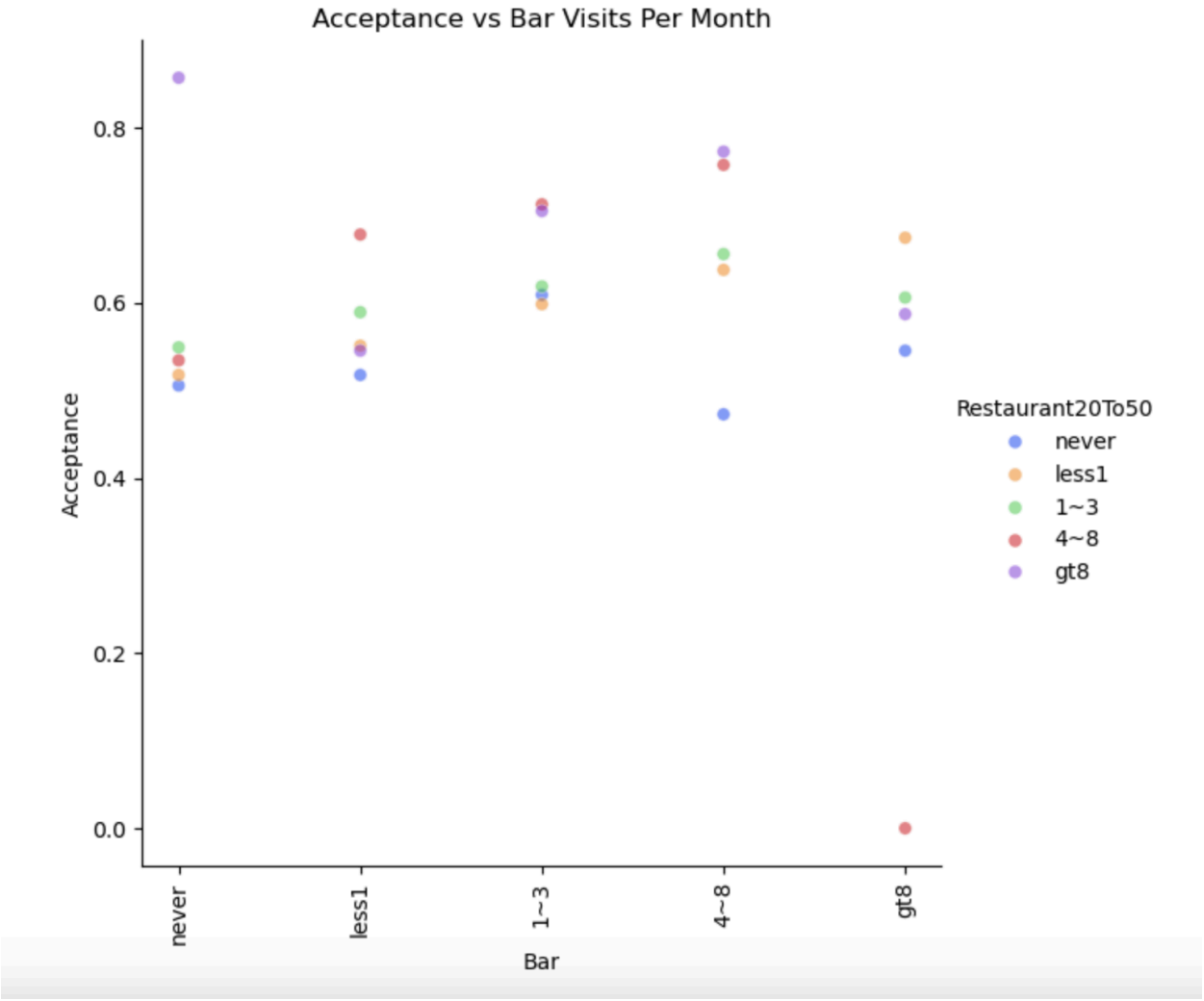
- Although both the lowest and highest income brackets have the same average acceptance rate across all ages, the dispersion of the data is less for the higher income bracket
- Young adults (21 years old) above the 87.5K income mark, have a similar acceptance rate as adults who frequent bars more than 4 times a month.

I then generated another scatter plot of Acceptance vs Income but this time added the additional dimension of number of visits to a bar per month



A stronger relationship between the number of times one visits a bar per can be seen (> 4 per month) compared to income and coupon acceptance, supporting the conclusion drawn during the guided exercise portion of this module.

The final graph I generated included looking at Bar visits per month vs Acceptance with the added dimension of dining patterns per month including ‘CarryAway’, ‘‘RestaurantLessThan20’ and 'Restaurant20To50’. The narrowest band of the Bar vs Restaurant Cost and acceptance was with individuals who prefer dining out at the $20 to $50 range meaning that there is likely a high degree of overlap of frequent visitations to bars and more expensive restaurants in this demographic.


Acceptance vs Bar Visits Per Month
Acceptance
0.0
0.2
0.4
0.6
0.8
never
less1
1~3
4~8
gt8
Bar
Restaurant20To50
never
less1
1~3
4~8
gt8