Will a Customer Accept the Coupon?

A brief report by Thaier Hayajneh

Context

Imagine driving through town and a coupon is delivered to your cell phone for a restaraunt near where you are driving. Would you accept that coupon and take a short detour to the restaraunt? Would you accept the coupon but use it on a sunbsequent trip? Would you ignore the coupon entirely? 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?

How would you determine whether a driver is likely to accept a coupon?

.

Problems

1. **Read in the `coupons.csv` file.**

   data = pd.read_csv('data/coupons.csv')

➔ **2. Investigate the dataset for missing or problematic data.**

2. Investigate the dataset for missing or problematic data.

In [8]:
```
# Explore the missing data
missing_vals = None

### BEGIN SOLUTION
missing_vals = data.isnull().sum()
### END SOLUTION

# Answer check
print(type(missing_vals))
missing_vals
```

```
<class 'pandas.core.series.Series'>
```

Out[8]:
```
destination                0
passanger                  0
weather                    0
temperature                0
time                       0
coupon                     0
expiration                 0
gender                     0
age                        0
maritalStatus              0
has_children               0
education                  0
occupation                 0
income                     0
car                    12576
Bar                      107
CoffeeHouse              217
CarryAway                151
RestaurantLessThan20     130
Restaurant20To50         189
toCoupon_GEQ5min           0
toCoupon_GEQ15min          0
toCoupon_GEQ25min          0
direction_same             0
direction_opp              0
Y                          0
dtype: int64
```

3. Decide what to do about your missing data -- drop, replace, other...

```python
# Investigate missing data in specified columns
missing_data = data[['Bar', 'CoffeeHouse', 'CarryAway', 'RestaurantLessThan20', 'Restaurant20To50']].isnull().sum()
print("Missing data in specified columns:")
print(missing_data)

# Drop missing data in specified columns
data.dropna(subset=['Bar', 'CoffeeHouse', 'CarryAway', 'RestaurantLessThan20', 'Restaurant20To50'], inplace=True)

# Confirm the removal of missing data
print("Number of rows after dropping missing data:", data.shape[0])
```

```
Number of rows: 12079
Number of columns: 26
Missing data in specified columns:
Bar                     0
CoffeeHouse             0
CarryAway               0
RestaurantLessThan20    0
Restaurant20To50        0
dtype: int64
Number of rows after dropping missing data: 12079
```

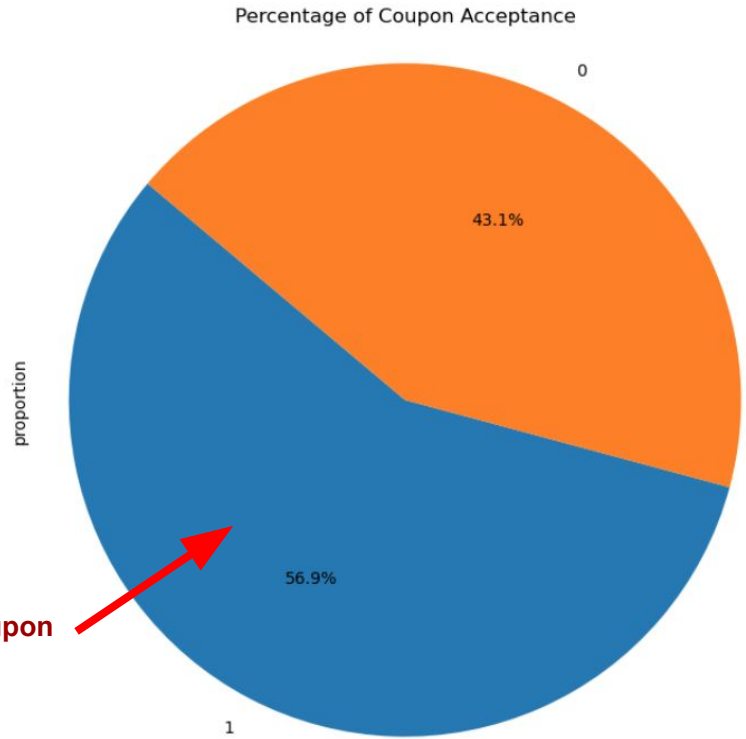
```
#  missing data in specified columns were dropped
# Now will do some cleaning and renaming to improve the readability of the data

# Rename the column 'Y' to 'Coupon accep'
data.rename(columns={'Y': 'Coupon accep'}, inplace=True)
```

In [78]:
```
# Now will do more cleaning an renaming to improve the readability of the data

# Replace 50plus with 50 in  the column 'age'
# Replace "plus" with "+" and below21 with 20

data['age'] = data['age'].str.replace('50+', '50')
data['age'] = data['age'].str.replace('below21', '20')
# Display the updated DataFrame
data.sample(12)

# This will make it easier to work with age as an integer
```

Out[78]:

	destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	...	CoffeeHouse	Ca
8174	No Urgent Place	Friend(s)	Snowy	30	10PM	Carry out & Take away	1d	Female	46	Married partner	...	4~8	
9672	No Urgent Place	Alone	Snowy	30	2PM	Bar	1d	Male	21	Single	...	never	
6450	Work	Alone	Sunny	55	7AM	Coffee House	1d	Female	26	Single	...	never	
7512	No Urgent Place	Friend(s)	Snowy	30	10AM	Carry out & Take away	1d	Female	41	Unmarried partner	...	1~3	
11429	Work	Alone	Snowy	30	7AM	Coffee House	1d	Male	31	Married partner	...	1~3	
8750	Home	Alone	Sunny	30	6PM	Carry out & Take away	2h	Female	41	Married partner	...	1~3	

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



56.93% Accepted the coupon

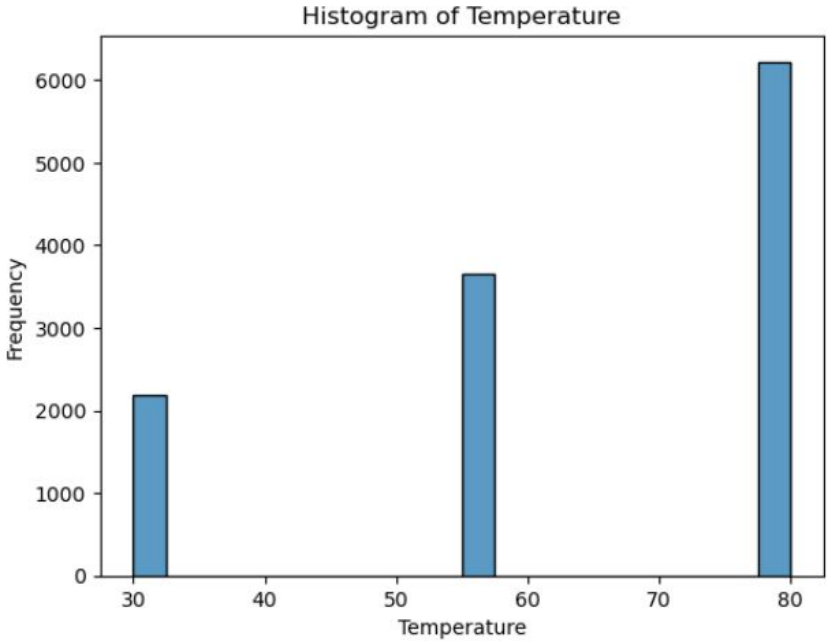
5. Use a bar plot to visualize the `coupon` column.

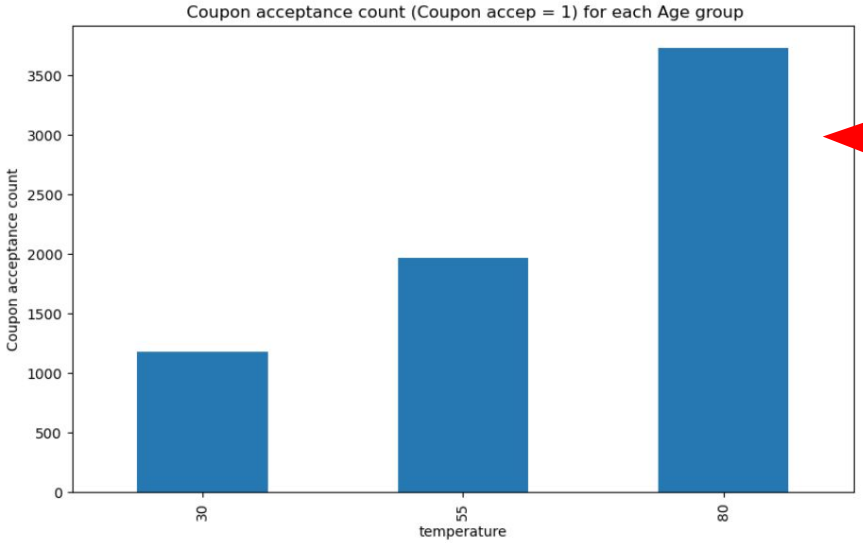
In [36]:
```
#Bar plot coupon accptance with Seaborn


sns.countplot(data = data, x = 'Coupon accep')

plt.xlabel('Coupon acceptance  [1 is Yes, 0 is No]')
plt.ylabel('Count')
plt.title('Count of Coupon Acceptance')
```

Out[36]:
```
Text(0.5, 1.0, 'Count of Coupon Acceptance')
```

6. Use a histogram to visualize the temperature column.

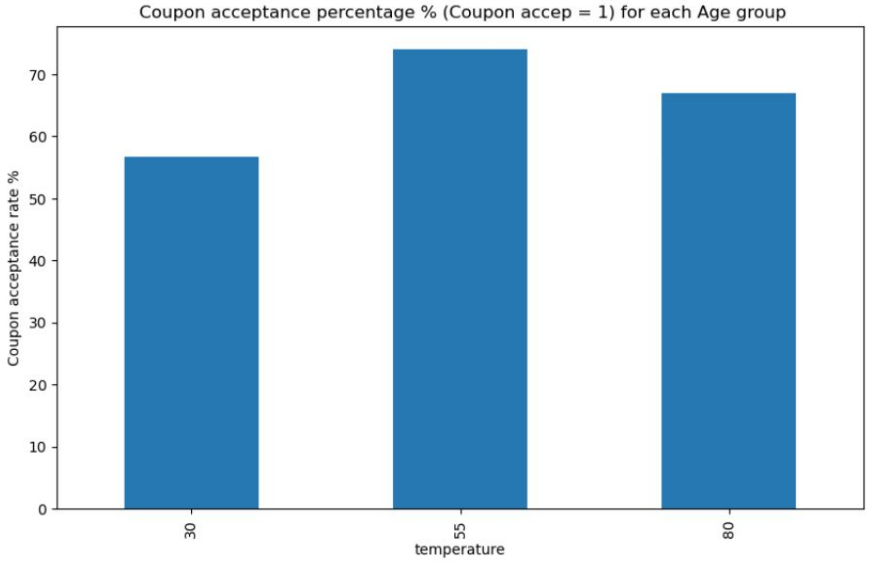
In [300]:
```
# Create the histogram
sns.histplot(data=data, x='temperature', bins=20)  # Adjust the number of bins as needed

# Add labels and title
plt.xlabel('Temperature')
plt.ylabel('Frequency')
plt.title('Histogram of Temperature')

# Show the plot
plt.show()
```





Significantly More drivers in warm weather accepted the coupon

Acceptance rate is higher in moderate weather



Investigating the Bar Coupons

Now, we will lead you through an exploration of just the bar related coupons.

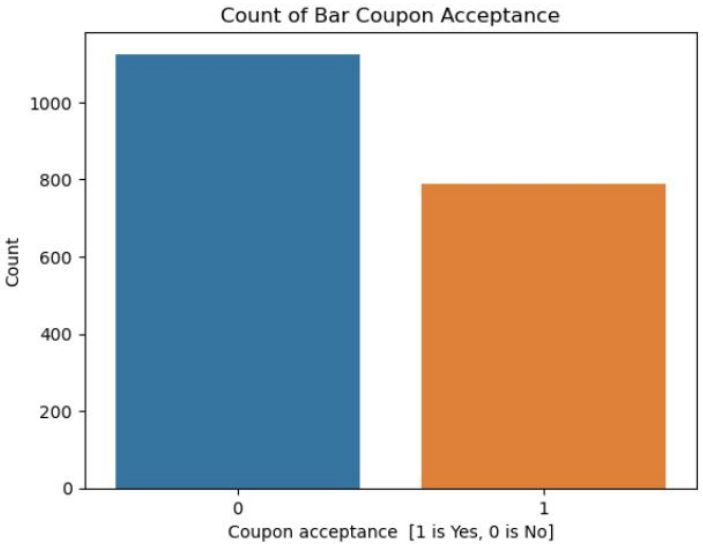
1. Create a new `DataFrame` that contains just the bar coupons.

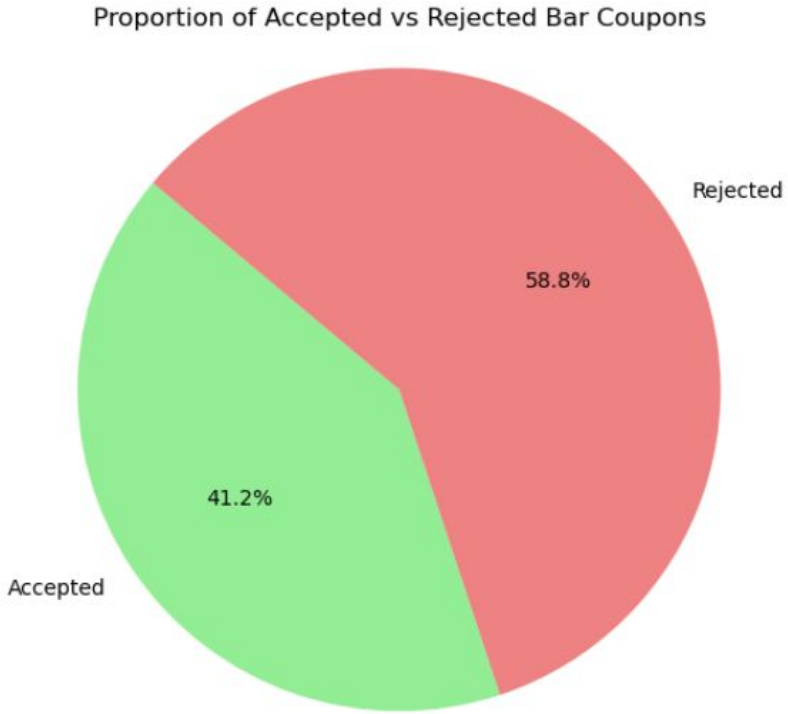
In [81]:
```
# Create a new DataFrame containing only rows where 'coupon' column is 'bar'
bar_coupons = data[data['coupon'] == "Bar"]

# Display the new DataFrame
bar_coupons.sample(12)
```

Out[81]:

	destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	...	CoffeeHouse	CarryAway	Restaur
5565	No Urgent Place	Alone	Sunny	80	10AM	Bar	1d	Male	36	Married partner	...	less1	less1	
6874	Home	Alone	Sunny	55	6PM	Bar	2h	Female	50	Single	...	4~8	1~3	
9542	Home	Alone	Rainy	55	6PM	Bar	1d	Female	20	Unmarried partner	...	1~3	less1	
1189	Work	Alone	Sunny	55	7AM	Bar	1d	Male	26	Married partner	...	gt8	4~8	

2. What proportion of bar coupons were accepted?



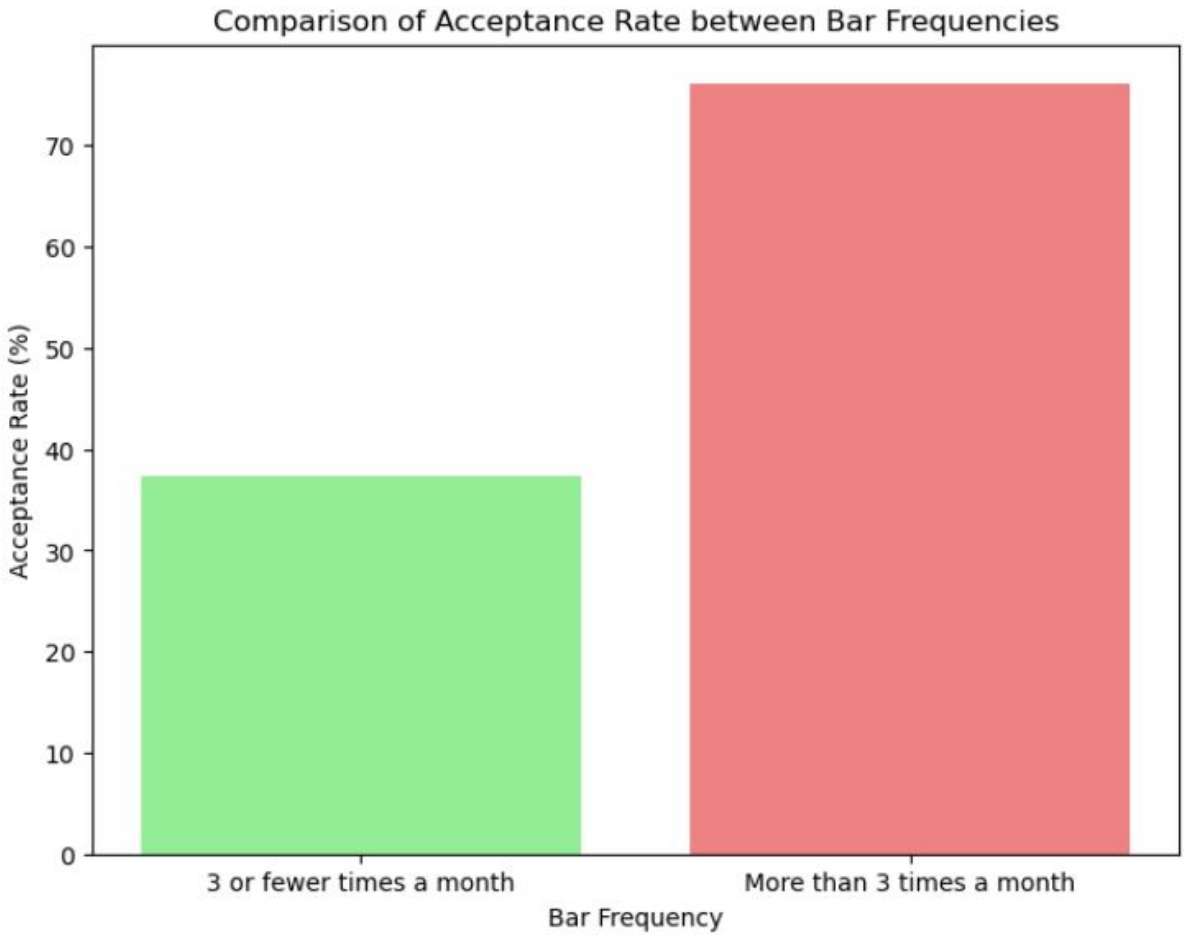


Total bar coupons: 1913
Total accepted bar coupons: 788
Proportion of accepted bar coupons:
0.41191845269210664

Bar coupon overall acceptance is 41% which means it is less likely for drivers to accept it

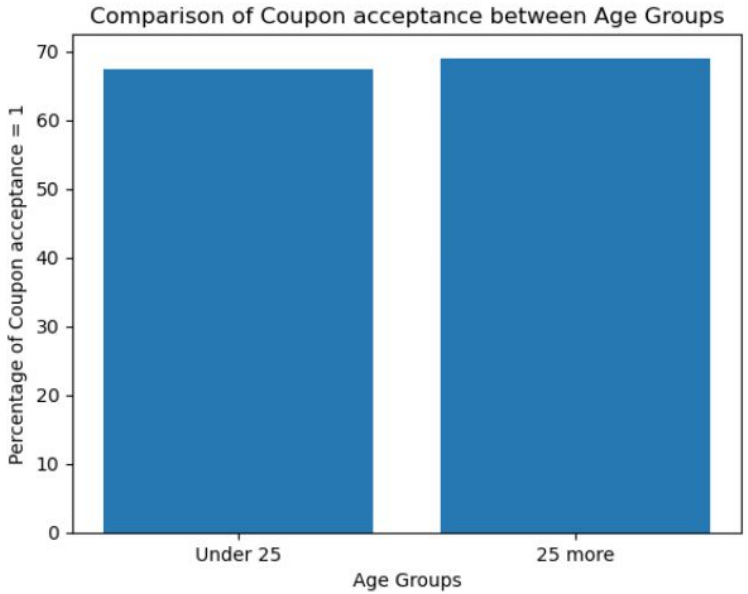
3. Compare the acceptance rate between those who went to a bar 3 or fewer times a month to those who went more.

I used groupby and size to determine the acceptance rate of drivers in each Bar group



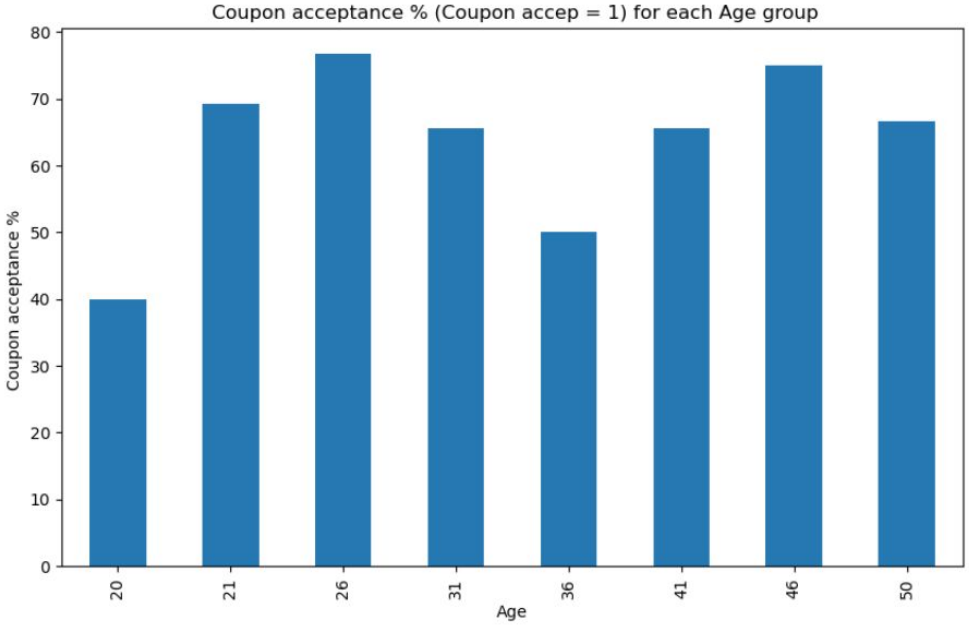
Observation:
Acceptance rate of those who went to a bar 3 or fewer times a month is **~37.3%** as compare to **~76.2%** for those who went more than 3 times, which makes sense as drivers who goes to the Bar more often are more likely to accept the Bar coupon

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?



Acceptance rate between drivers who go to a bar more than once a month and are under the age of 25: 67.45562130177515
Acceptance rate between drivers who go to a bar more than once a month and are over the age of 25: 68.98263027295285

I Calculated the acceptance rate for each 'age' value, just to ensure the result from the previous diagram are correct



Observation: the comparison between the acceptance rate of drivers who go to a bar more than once a month and are over the age of 25 to all others indicates no significant difference. Both groups exhibit a high acceptance rate of over 2/3, suggesting that regardless of age, drivers who go to bars frequently tend to accept bar coupons at a similar rate. Therefore, age does not appear to significantly impact the acceptance rate among drivers who visit bars frequently.

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.

Acceptance rate between drivers who go to a bar more than once a month and without a kid: 70.943

Acceptance rate between drivers who go to a bar more than once a month and without a kid and had occupations other than farming, fishing, or forestry.: 70.943

Observation: The acceptance rate among drivers who go to a bar more than once a month and do not have a child is 71%, indicating a high acceptance rate, which is expected given their frequent bar visits. As for drivers with occupations other than farming, fishing, or forestry, there are none represented in the dataset of bar_more1 for drivers who go to bars more than once a month , so there is no change in their acceptance rate.

6. Compare the acceptance rates between those drivers who:

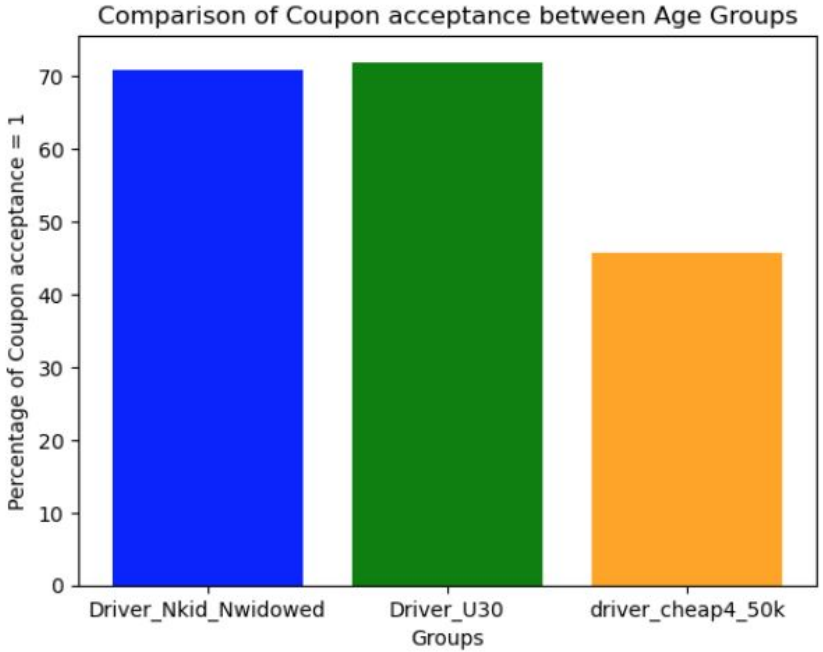
- go to bars more than once a month, had passengers that were not a kid, and were not widowed *OR*
- go to bars more than once a month and are under the age of 30 *OR*
- go to cheap restaurants more than 4 times a month and income is less than 50K.

Acceptance rate between drivers who go to a bar more than once a month and without a kid as a passenger and were not widowed: 70.94339622641509

Acceptance rate between drivers who go to a bar more than once a month and Under the age of 30 is: 71.95121951219512

Acceptance rate between drivers who not necessarly go to a bar more than once a month and go to cheap restaurants more than 4 times a month and has income is less than 50K: 45.645645645645644

Observation: The acceptance rate among drivers who go to a bar more than once a month and do not have a child is 71%, indicating a high acceptance rate, which is expected given their frequent bar visits. As for drivers with occupations other than farming, fishing, or forestry, there are none represented in the dataset of bar_more1 for drivers who go to bars more than once a month , so there is no change in their acceptance rate.



Observation:

- The acceptance rate among drivers who go to a bar more than once a month and do not have a child as a passenger, and are not widowed, is approximately 70.94%. This suggests that this group of drivers tends to have a relatively high acceptance rate for bar coupons.
- The acceptance rate among drivers who go to a bar more than once a month and are under the age of 30 is approximately 71.95%. This indicates that younger drivers who frequent bars frequently also exhibit a high acceptance rate for bar coupons.
- In contrast, the acceptance rate among drivers who may not necessarily go to a bar more than once a month but visit cheap restaurants more than 4 times a month and have an income of less than 50K is approximately 45.65%. This suggests that this group of drivers has a lower acceptance rate for bar coupons compared to the other two groups.
- Overall, the results indicate variations in acceptance rates based on different demographic and behavioral factors among drivers.

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

- Frequency of Bar Visits: Drivers who go to a bar more than three times a month have a significantly higher acceptance rate for bar coupons (~76.2%) compared to those who go three or fewer times (~37.3%). This suggests that frequent bar visitors are more likely to accept bar coupons, which aligns with the expectation that individuals who frequent bars may be more inclined to take advantage of such offers.
- Age Factor: The comparison between drivers under the age of 25 and those over the age of 25 who go to a bar more than once a month indicates no significant difference in acceptance rates. Both groups exhibit high acceptance rates of over 2/3 (~67.5% and ~69%, respectively). This suggests that age does not play a significant role in determining the acceptance rate among drivers who visit bars frequently.
- Child Passengers and Marital Status: Drivers who go to a bar more than once a month and do not have a child as a passenger, and are not widowed, demonstrate a relatively high acceptance rate for bar coupons (~71%). This indicates that this demographic segment tends to be more receptive to bar coupons, possibly due to their lifestyle choices or disposable income.
- Behavioral Patterns: On the other hand, drivers who may not necessarily go to a bar more than once a month but visit cheap restaurants more than four times a month and have an income of less than 50K show a lower acceptance rate for bar coupons (~45.6%). This suggests that behavioral patterns, such as dining preferences and income level, can influence the acceptance of bar coupons.
- Overall, these observations highlight the variability in acceptance rates among drivers based on demographic factors, frequency of bar visits, and behavioral patterns. Factors such as age, presence of child passengers, marital status, and income level can all contribute to differences in acceptance rates among drivers.

7. Independent Investigation

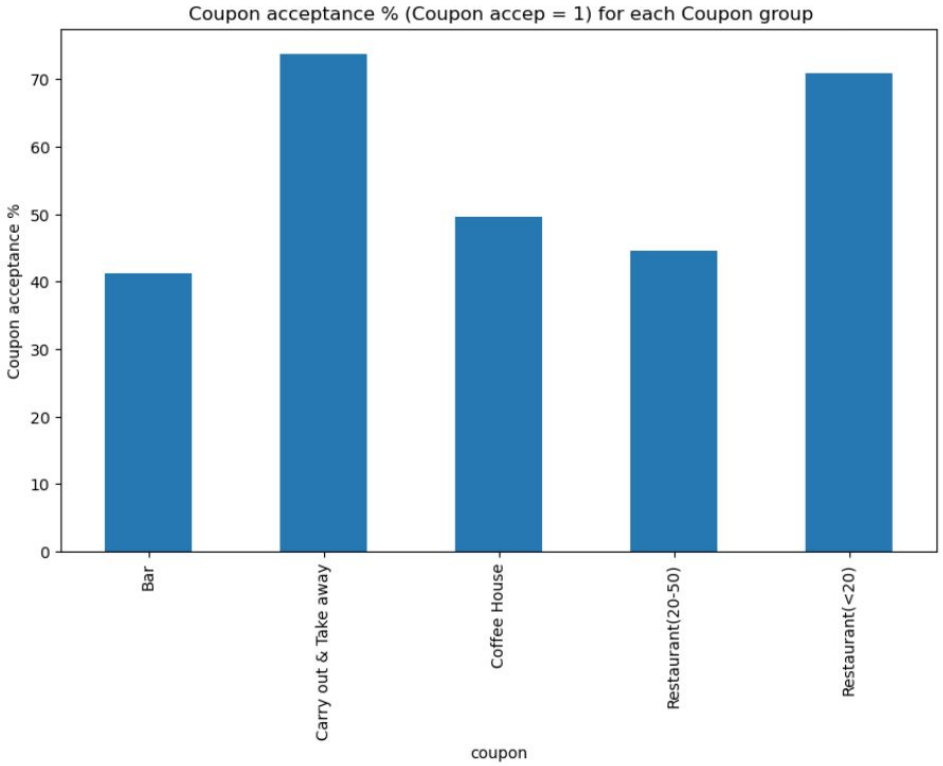
Using the bar coupon example as motivation, you are to explore one of the other coupon groups and try to determine the characteristics of passengers who accept the coupons.

I decided to choose

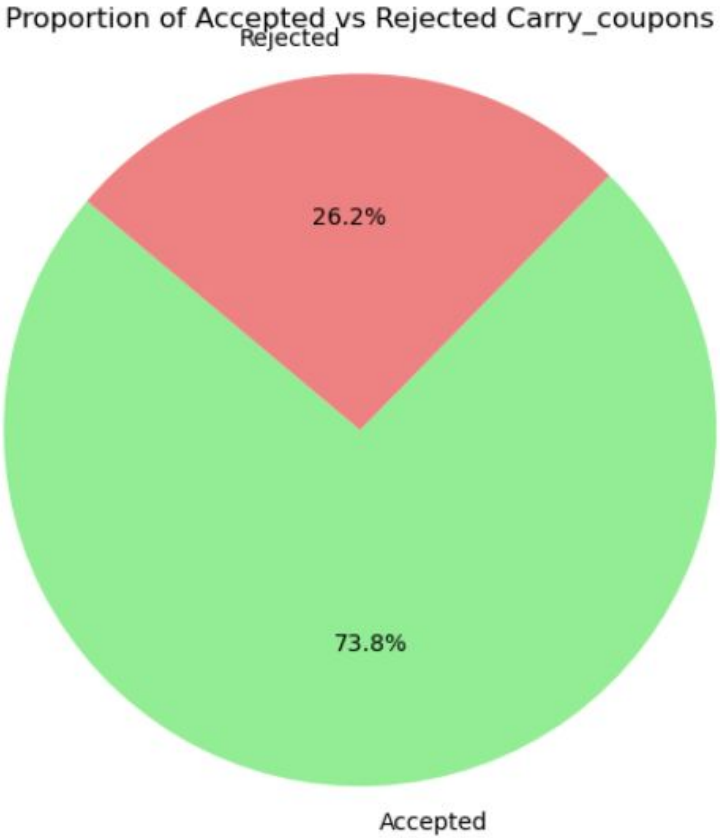
"Carry out & Take away"

'coupon' to do my analysis as similar to the Bar analysis.

It has the highest acceptance ratio



"Carry out & Take away" → will refer to as “Carry”




```
## Columns that can be explored
print(Carry_coupons.columns)
```

```
Index(['destination', 'passanger', 'weather', 'temperature', 'time', 'coupon',
       'expiration', 'gender', 'age', 'maritalStatus', 'has_children',
       'education', 'occupation', 'income', 'car', 'Bar', 'CoffeeHouse',
       'CarryAway', 'RestaurantLessThan20', 'Restaurant20To50',
       'toCoupon_GEQ5min', 'toCoupon_GEQ15min',
'toCoupon_GEQ25min',
       'direction_same', 'direction_opp', 'Coupon accep'],
      dtype='object')
```

To explore the acceptance of Carry coupon group, I followed these steps:

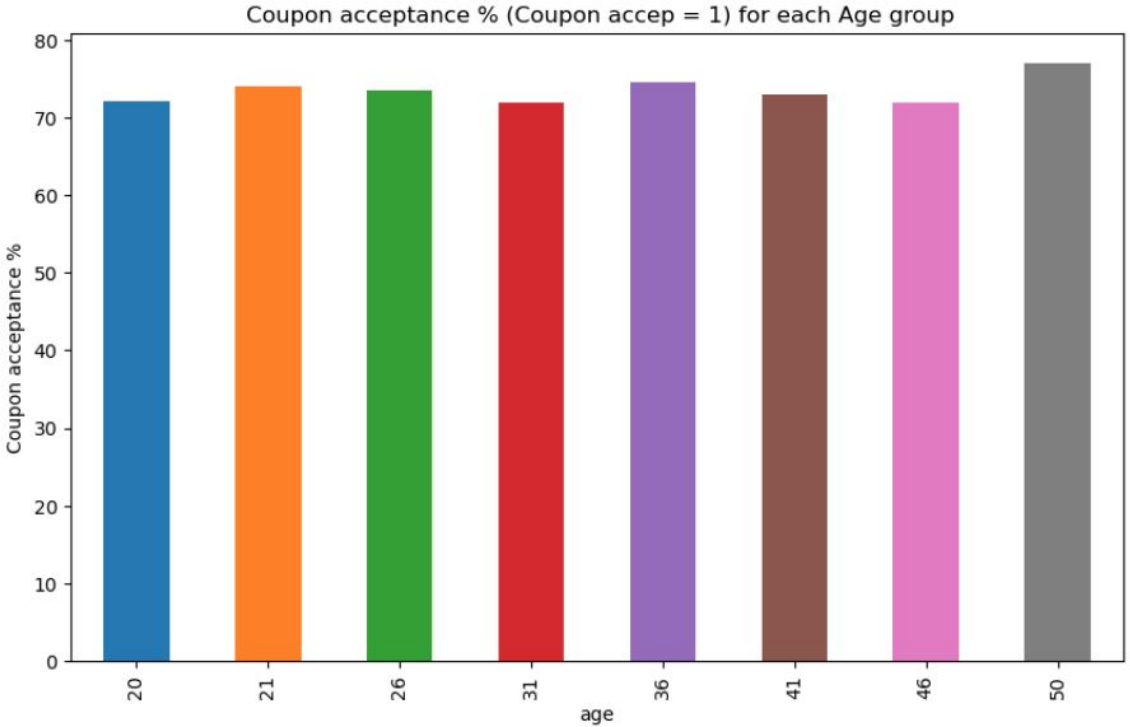
1. **Data Understanding:** Familiarize myself with the dataset and understand the attributes available. I have already provided a detailed description of the attributes as a starting point
2. **Initial Data Exploration**: Begin by exploring the distribution of the target variable ('Y' or 'Coupon accep') I used visualizations such as bar plots, histograms, to compare acceptance rates across different categories.
3. **Multivariate Analysis**: I Explored interactions between multiple features and how they influence coupon acceptance. For example, I will analyze how the acceptance rate varies based on combinations of gender, age, marital status, number of children, etc.
4. **Correlation Analysis**: Investigate correlations between features and the target variable. This can help identify which features have a stronger influence on coupon acceptance.
5. **Temporal Analysis**: Explore how acceptance rates vary with time of day, day of the week, or other temporal factors. You can use line plots or bar plots to visualize these trends.
6. **Spatial Analysis**: If geographical data is available, analyze how location influences coupon acceptance.
7. **Weather and Temperature Impact**: Investigate the impact of weather conditions and temperature on coupon acceptance.

I started by exploring how the various feature will impact the acceptance of the "Carry out & Take away" coupon

- I used groupby and bar diagrams
- In this presentation I will only show the code and plot for the 'age' and then report the observations for all the other variables

In [319]:
```
# exmplore the impat of age on carry out
#percentage_accepted = selected_groups[selected_groups['Coupon accep'] == 1].groupby('age').size() / selected_groups

percentage_accepted = Carry_coupons[Carry_coupons['Coupon accep'] == 1].groupby('age').size() / Carry_coupons.groupb

# Plot the count as a bar diagram
percentage_accepted.plot(kind='bar', figsize=(10, 6),color=plt.cm.tab10.colors)

# Add labels and title
plt.xlabel('age')
plt.ylabel('Coupon acceptance count')
plt.title(' Coupon acceptance count (Coupon accep = 1) for each Age group')

# Show the plot
print(percentage_accepted)
print(type(percentage_accepted))
plt.show()
```

```
age
20    72.115385
21    74.034335
26    73.469388
31    71.891892
36    74.476987
41    73.039216
46    71.900826
50    77.014925
dtype: float64
<class 'pandas.core.series.Series'>
```

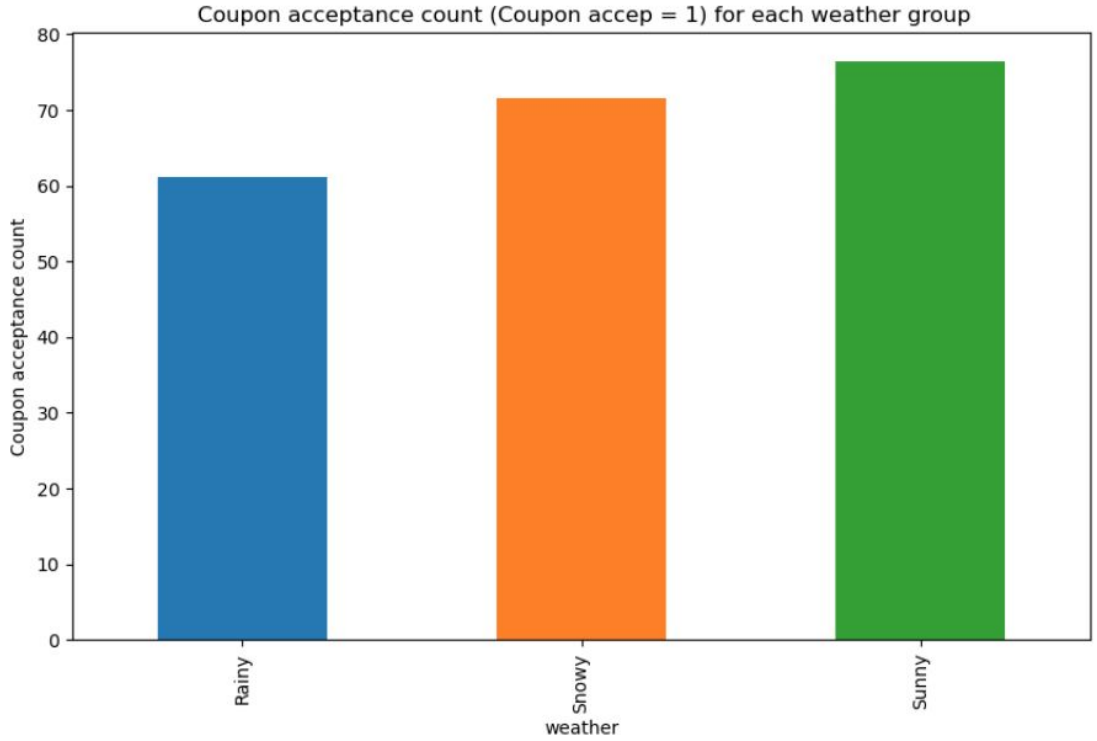


Observation on ‘age’: Acceptance rate for "Carry out & Take away" coupon is high for all ages...

However, it is a bit higher for ages 50 and above, this is expected as these age group are more likely to order take out.

Results of exploring impact of other variable on the acceptance of Carry coupon:

1. **Impact of weather on carry out:** (Rainy: 61.1%, Snowy: 71.6, Sunny: 76.40)
Surprisingly, carry out coupons acceptance is higher when the weather is not rainy or snowy



Results of exploring impact of other variable on the acceptance of Carry coupon:

2. Impact of carry out frequency history:

- For drivers who order takeout:
  - More than 8 times a week: Acceptance rate is 75.34%
  - 4 to 8 times a week: Acceptance rate is 75.35%
  - 1 to 3 times a week: Acceptance rate is 74%
  - One or less than once a week: Acceptance rate is 67.9%
  - Never ordered carry out: Acceptance rate is 78.6%

The trend is as expected: drivers who usually order carry out more often are more likely to accept the coupon. However, there is an interesting exception or anomaly in the results. Drivers who never ordered carry out have the highest acceptance rate. Upon further exploration, it was discovered that the sample size for this group is significantly smaller compared to the other groups, with only 28 drivers, out of which 22 accepted the coupon. This small sample size could skew the results, and it is also possible that these 22 drivers were curious to try the coupon and decided to order carry out for the first time. Further investigation may be warranted to better understand this anomaly.

CarryAway group	More than 8	4 to 8	1 to 3	1 or less	never
Total received coupon	296	783	830	343	28
Accepted the coupon	233	590	614	233	22

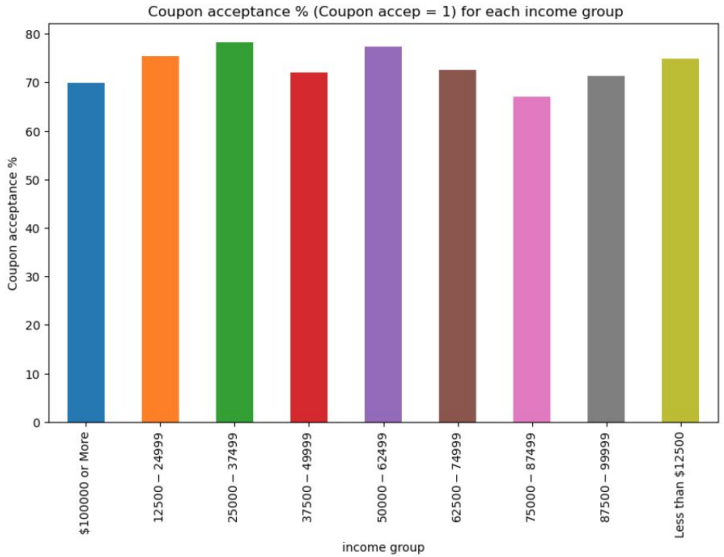
Results of exploring impact of other variable on the acceptance of Carry coupon:

3. Impact of having children:

- For drivers who have children or not the acceptance rate did not noticeably changed.

4. Impact of income :

- It was noticed that income groups [$25,000 - $37,000] (78.3%) and [$50,000 - $62,500](77.4%) has the highest acceptance rate
- Overall, the higher the income the less acceptance for carry out coupon



Results of exploring impact of other variable on the acceptance of Carry coupon:

5. Impact of driving direction:

- For drivers who are going in the same direction of the carry out coupon invitation the acceptance rate was 75.4% and for those going not going in same direction it was 70.6%.
- This is not expected but it also implies that coupon acceptance decision is not impacted with the driving direction.

6. Impact of number of times drivers going to Restaurants that cost Less Than $20 :

- For drivers who [go to Restaurants that cost Less Than $20] more frequently or less frequently the acceptance rate did not noticeably changed.

7. Impact of CoffeeHouse frequency history:

- For drivers who go to CoffeeHouse:
  - More than 8 times a week: Acceptance rate is 60.7%
  - 4 to 8 times a week: Acceptance rate is 73.2%
  - 1 to 3 times a week: Acceptance rate is 74.7%
  - One or less than once a week: Acceptance rate is 75.6%
  - Never ordered carry out: Acceptance rate is 76.3%

The trend is as expected: drivers who usually go to CoffeeHouse more often are **LESS likely** to accept the coupon. It is obvious that they are more interested Coffee houses than carry out coupons.

Results of exploring impact of other variable on the acceptance of Carry coupon:

8. Impact of gender:

- For drivers who are male the acceptance rate was 76.2% and females it was 71.5%.
- This indicates that males are more likely to accept the carry out coupon than females..

9. Impact of Marital Status :

- For Single drivers the acceptance rate was 75.7% while (married/unmarried) partners has a lower acceptance rate (72.9%/71%)
- Obviously singles would prefer to carry out rather than dining in or cooking at home.
- Note that: I did not include data for widowed and divorced and their sample was significantly smaller.

```
maritalStatus
Divorced              78
Married partner      652
Single               677
Unmarried partner    258
Widowed               17
dtype: int64
```

Acceptance count

```
maritalStatus
Divorced             72.222222
Married partner      72.930649
Single               75.727069
Unmarried partner    70.879121
Widowed              85.000000
```

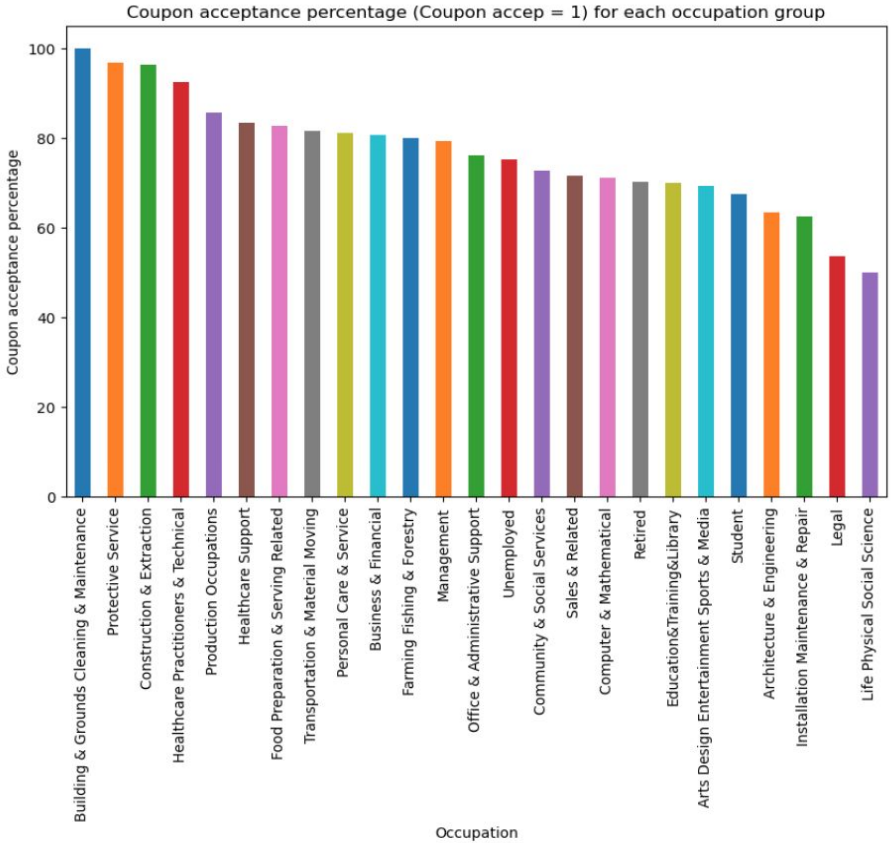
Acceptance %

Results of exploring impact of other variable on the acceptance of Carry coupon:

11. Impact of occupation frequency history:

Interesting finding; occupations:
"Building & Grounds Cleaning & Maintenance", "Protective Service" , "Construction & Extraction" , "Healthcare Practitioners & Technical" has the highest Carry on coupon acceptance rate %90+

On the other hand, the **lowest acceptance rate** is for occupations:
Arts Design Entertainment Sports & Media (69.3%);
Student (67.5%);
Architecture & Engineering (63.3%);
Installation Maintenance & Repair (62.5);
Legal (53.7%)
Life Physical Social Science (50%)



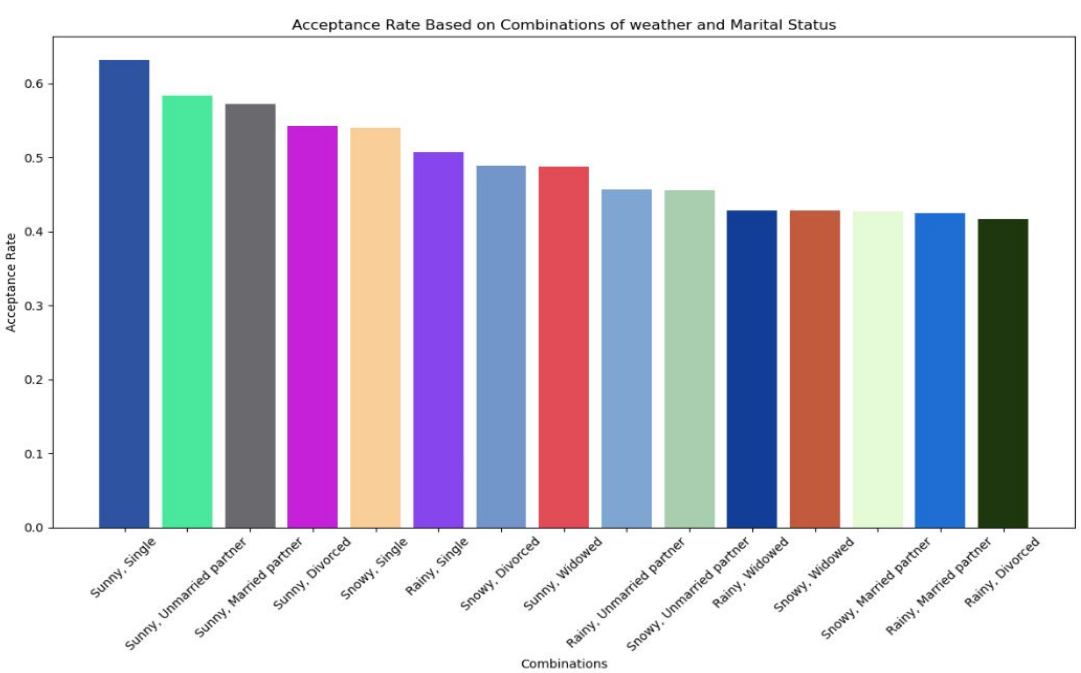
Results of exploring impact of other variable on the acceptance of Carry coupon:

12. Impact of both weather and marital status:

Interesting findings:

Acceptance rate is high in sunny days for singles and partners.

Acceptance rate is lower in snow and rain for partners



Results of exploring impact of other variable on the acceptance of Carry coupon:

13. Impact of both age and temperature:

Interesting findings:

Acceptance rate is higher in warm temperature for all ages.