COUPON USAGE ANALYSIS AND RECOMMENDATIONS

Byron Han

PROJECT BACKGROUND

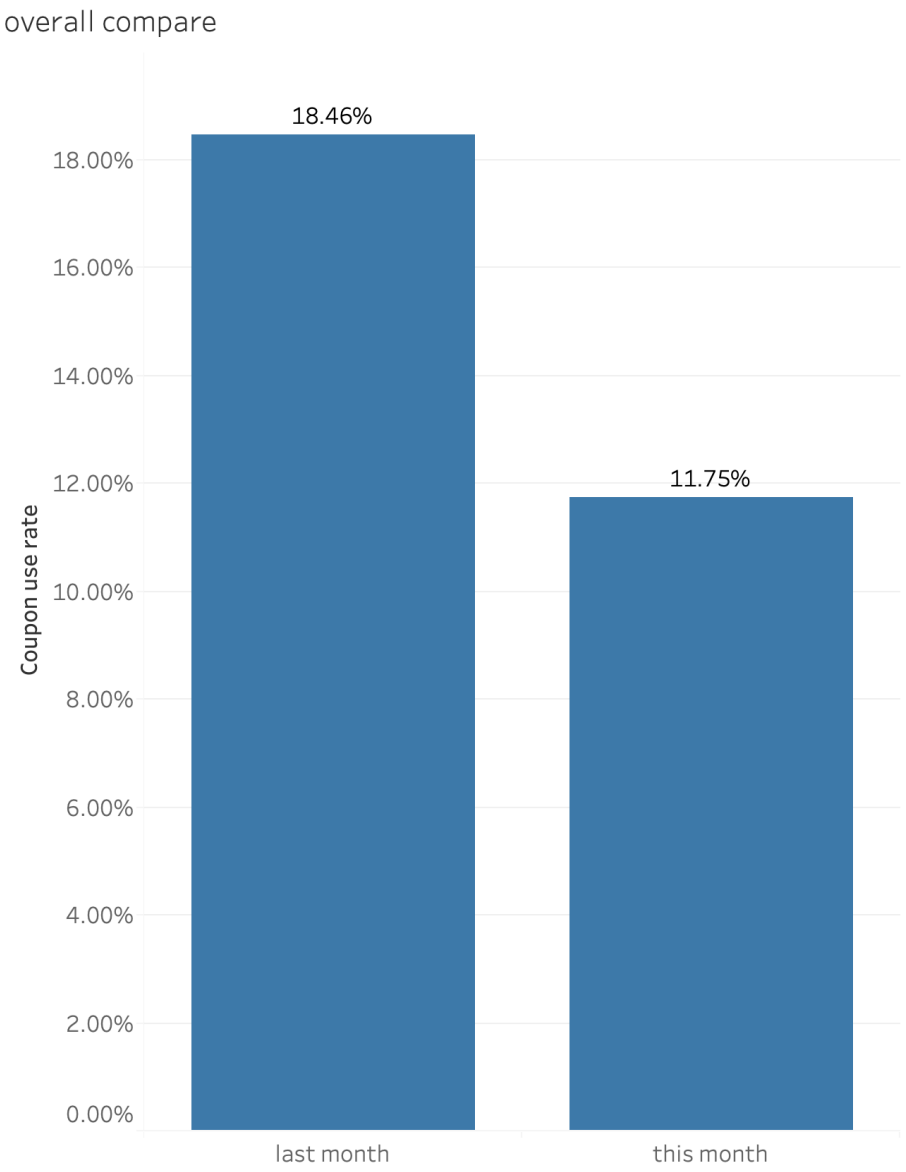
- A China-based fashion design company has been giving out coupons to its customers in the last 6 months.
- The operation team wants to evaluate the performance of coupons with its customers and receive further recommendations.
- Also the team wants to find out how to predict coupon usage.

QUESTIONS WILL BE ANSWERED IN THIS PRESENTATION

- Does the new coupon perform better than last month?
- What people are most likely to use new coupon?
- What people are most likely to use coupon in general?
- How should we plan coupon for next month?
- Can we predict whether coupon will be used?

SUMMARY OF FINDINGS AND RECOMMENDATIONS

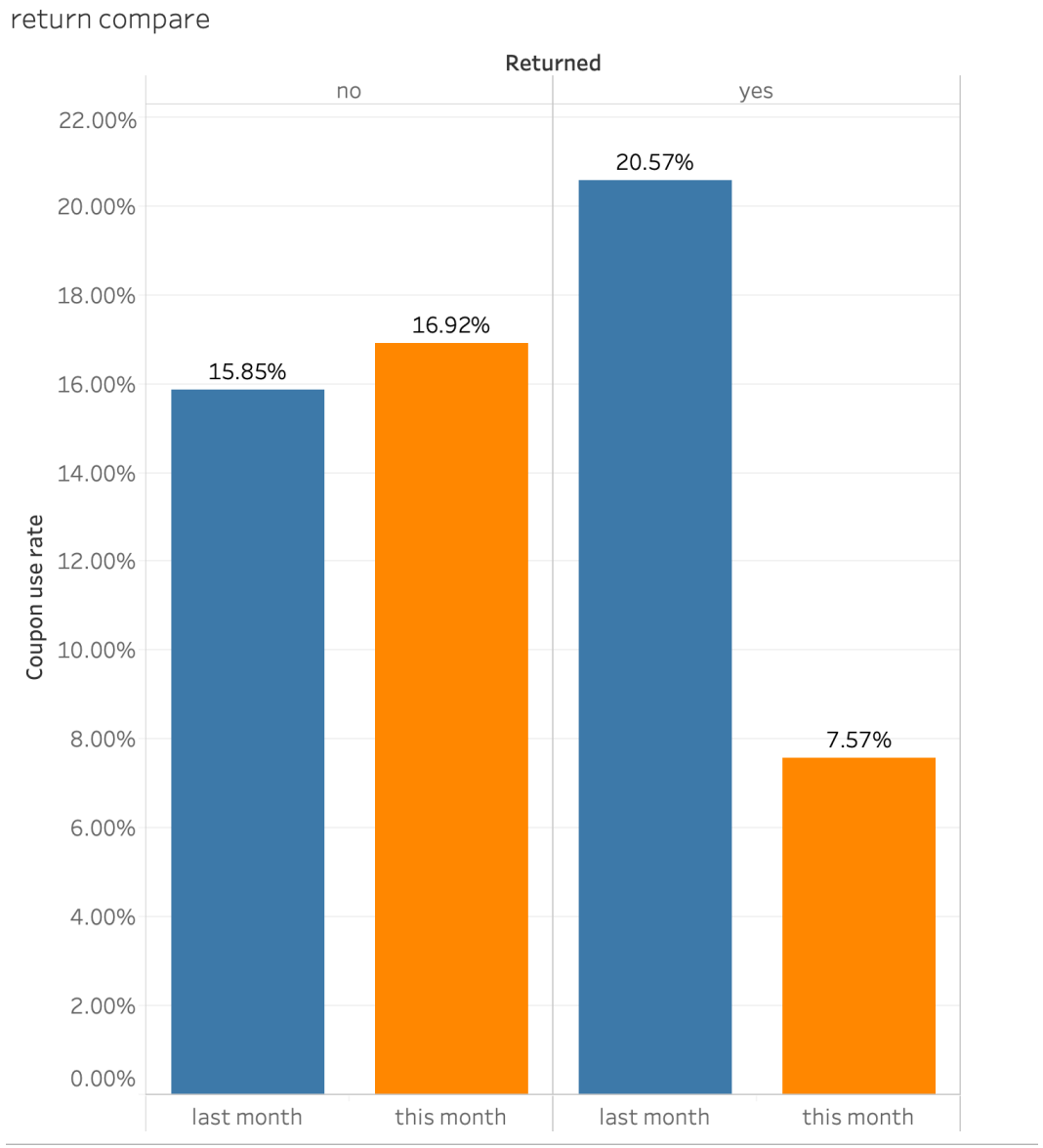
- New coupon is used less frequent than last month (-36%).
  - People returned items before are -66% less likely to use coupon.
  - Mid age group is -42% less likely to use coupon, where young and elder are +30% and +17% respectively.
- Overall, **non-working class** who has **never returned** before is more likely to use coupon this month.
- Re-evaluate if new coupon is biased on age and return policy
- Separate coupon strategy for management/blue-collar
- Accurate prediction is not achievable with current data source



HOW DOES NEW COUPON PERFORM COMPARE TO LAST MONTH?

- Overall coupon use rate dropped by ~36%.

- Down from 18.46% to 11.75%.

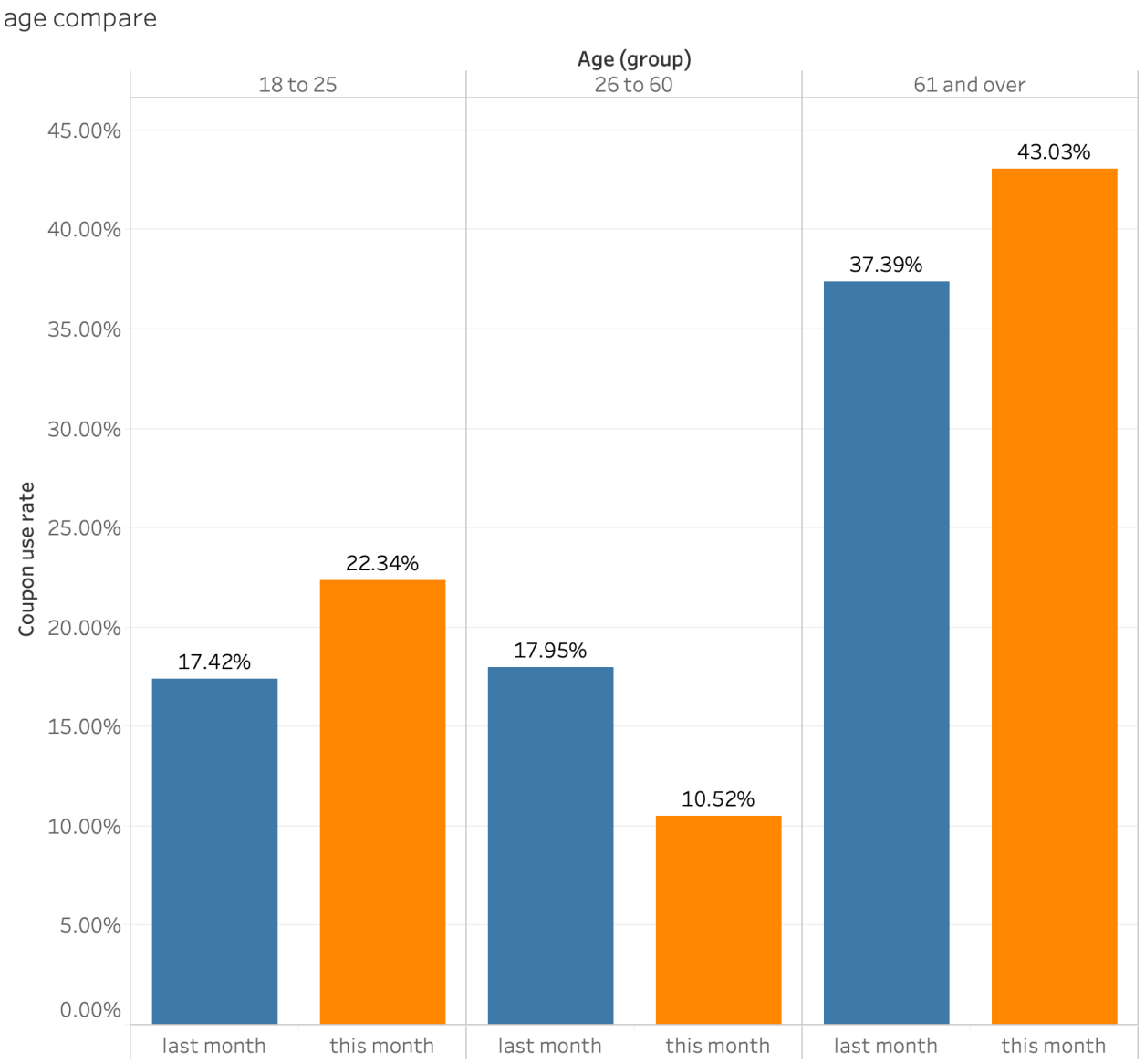


WHERE IS THE PROBLEM?

- While many variables don't show dramatic change, coupon use rate dropped by ~66% for those who have returned items before.

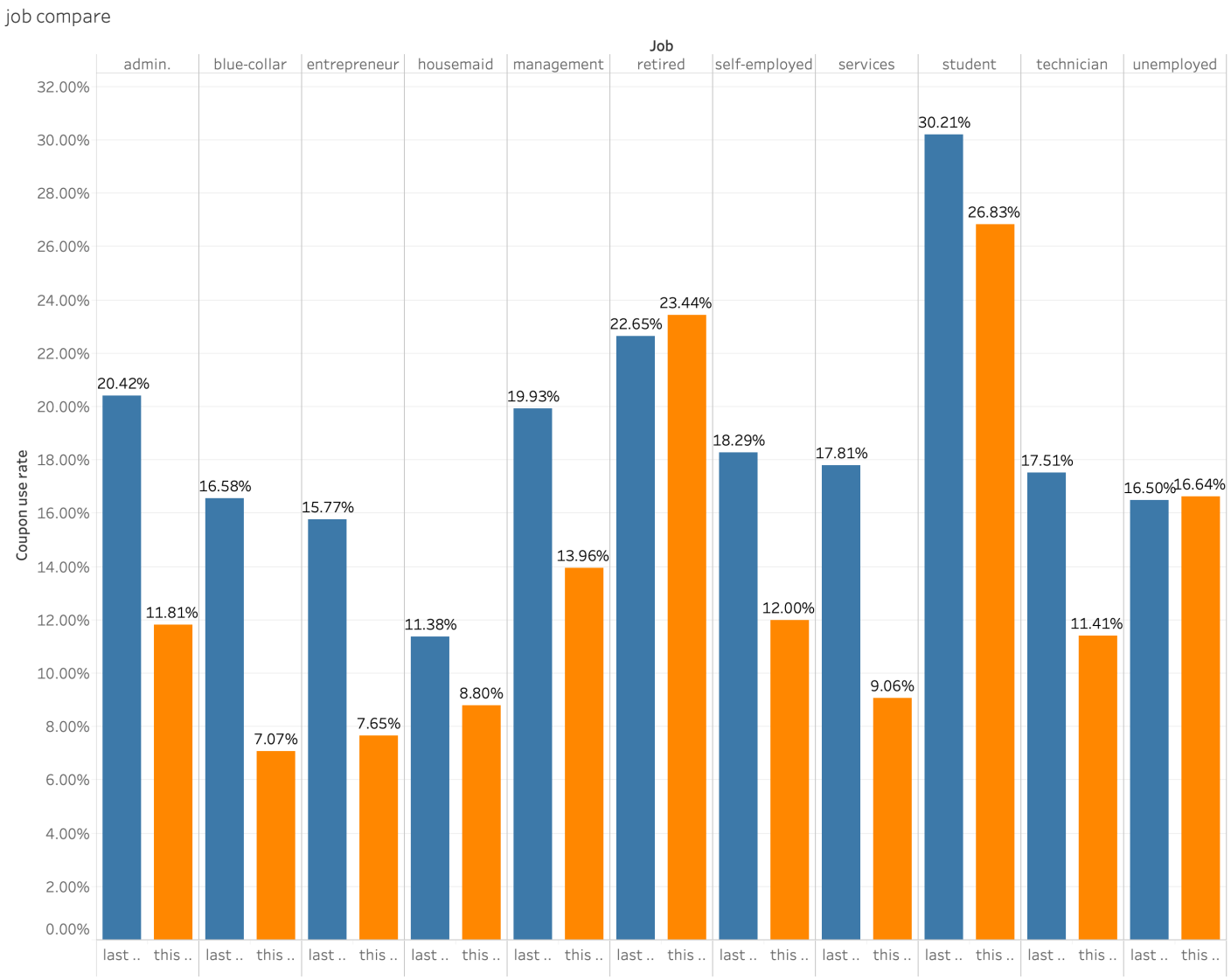
- The reason might be:

  - New coupon is biased towards people who has not returned anything before.
  - New coupon has stricter terms and condition on return policy
  - New coupon is only visible to those who has not returned anything before (if not anyone in the dataset can see coupon).



WHERE IS THE PROBLEM

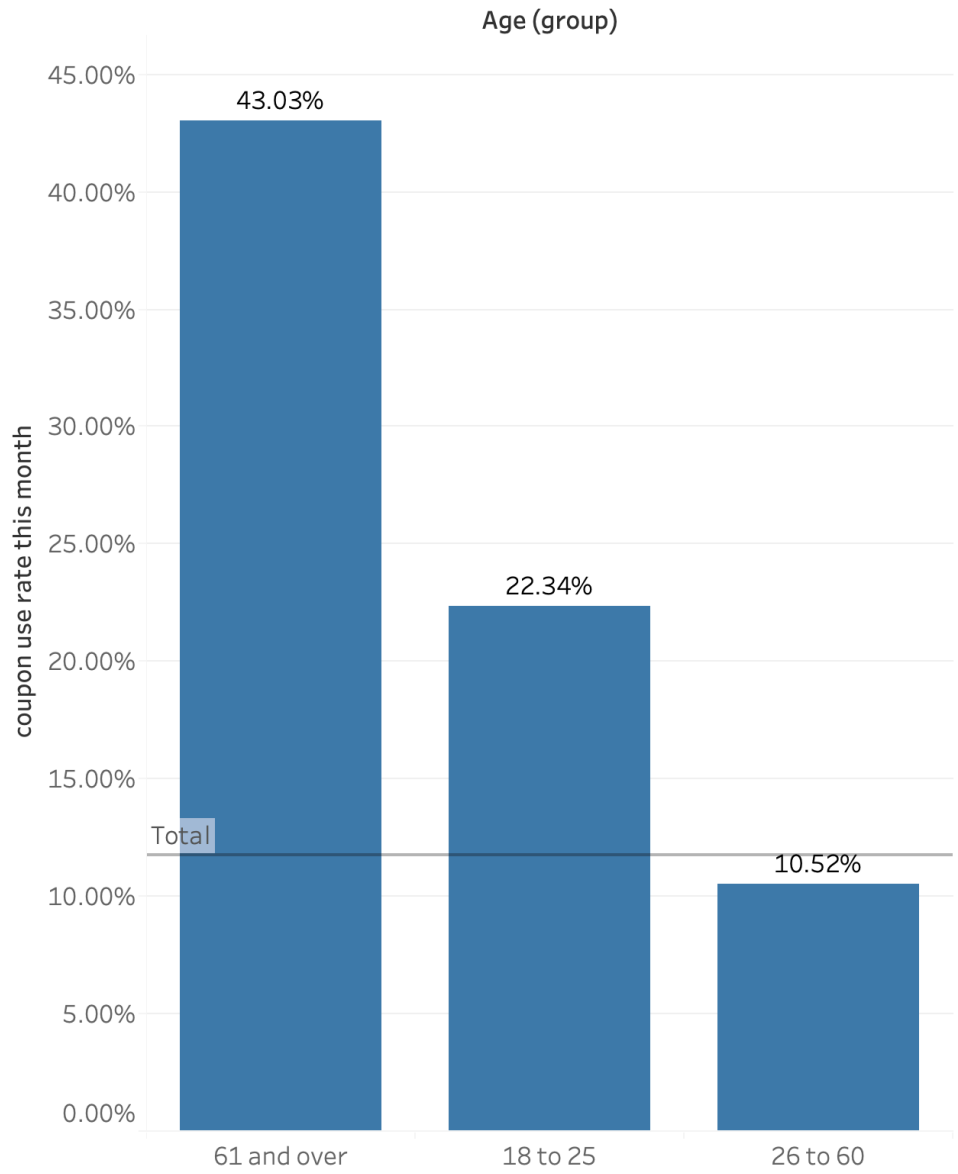
- Coupon use rate dropped by ~ 42% for mid age group.
- Whereas in younger and elder groups coupon are used more often.
- Not observable correlation between age group and return behavior (appendix 1).
- Therefore, it is likely coupon is designed bias towards mid age group.
- The reason might be:
  - Coupon is designed for certain age to use.
  - Applicable products are only attractive to certain age group.



WHERE IS THE PROBLEM

- Coupon use rate dropped in all job categories except “retired” and “unemployed”.
- Jobs is significantly correlated with age, but not noticeably correlated with return policy (appendix 2).
- Therefore, coupon is not likely designed bias towards jobs, but confoundedly affected by age.

age

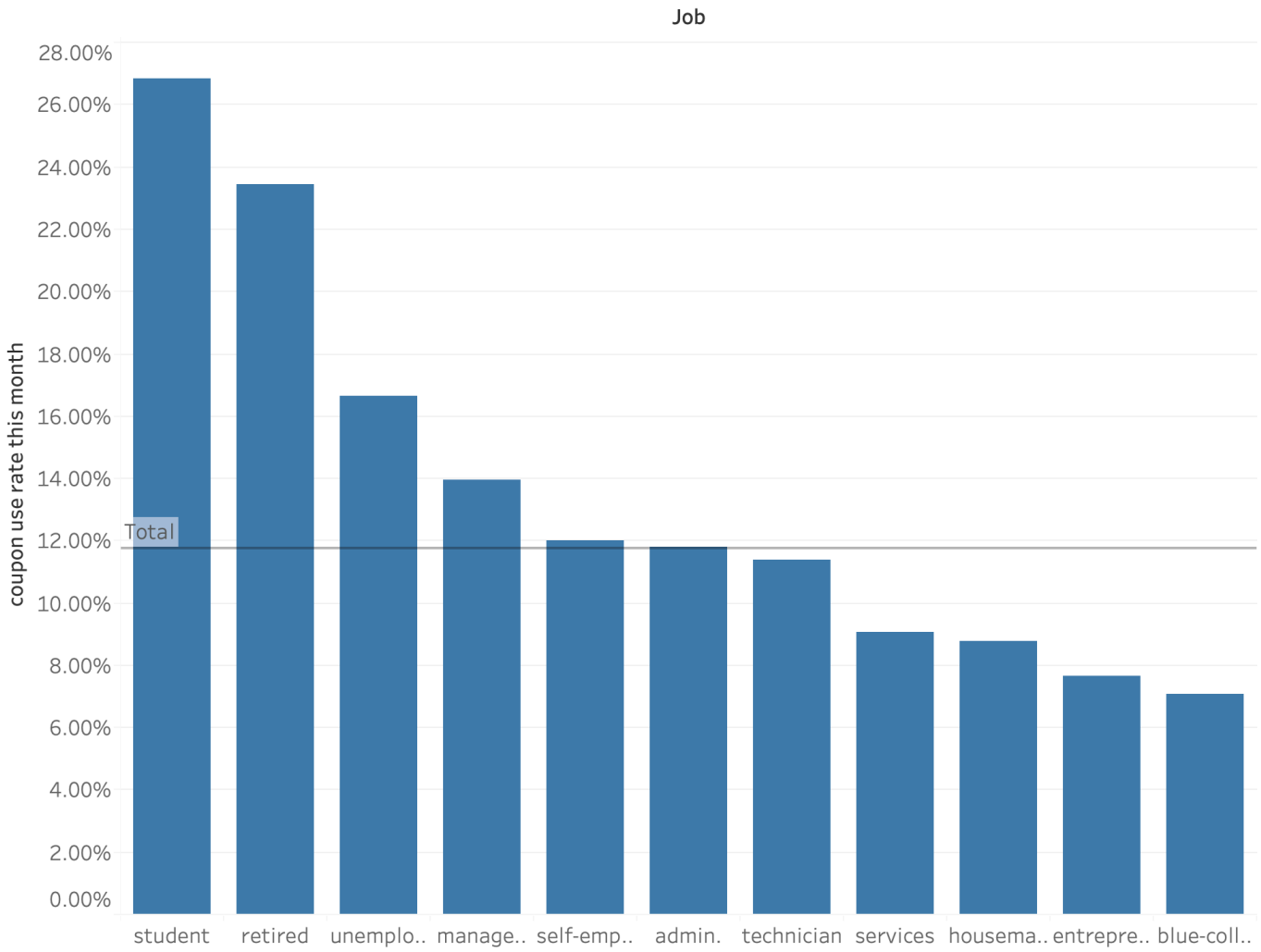


WHAT PEOPLE ARE MOST LIKELY TO USE NEW COUPON

Among all age, 60 years older people is most likely to use new coupon (+258%).

25-60 is least likely to use new coupon.

job

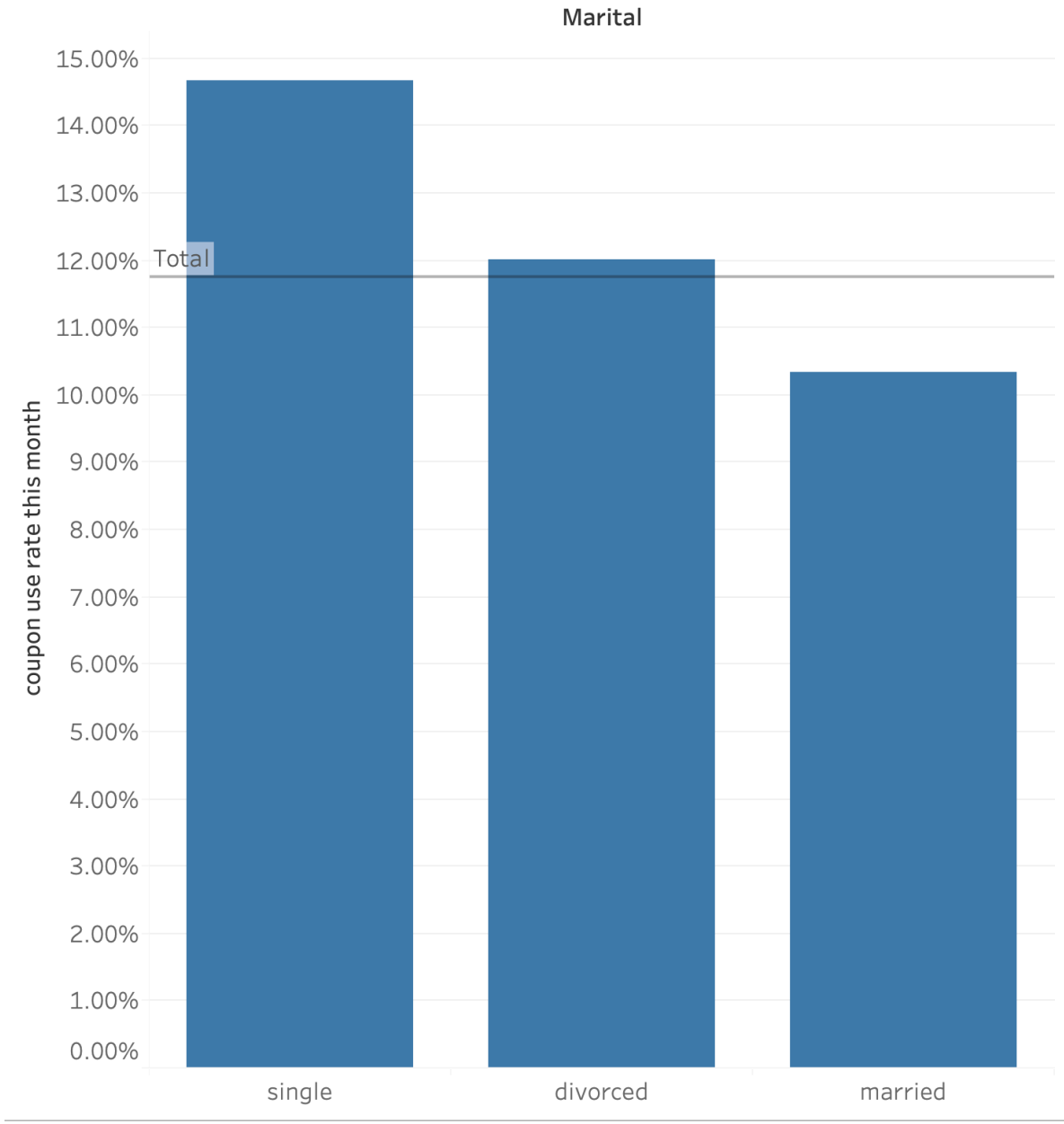


WHAT PEOPLE ARE MOST LIKELY TO USE NEW COUPON

Among all jobs, student is most likely to use new coupon (+133%).

Blue collar is least likely to use new coupon.

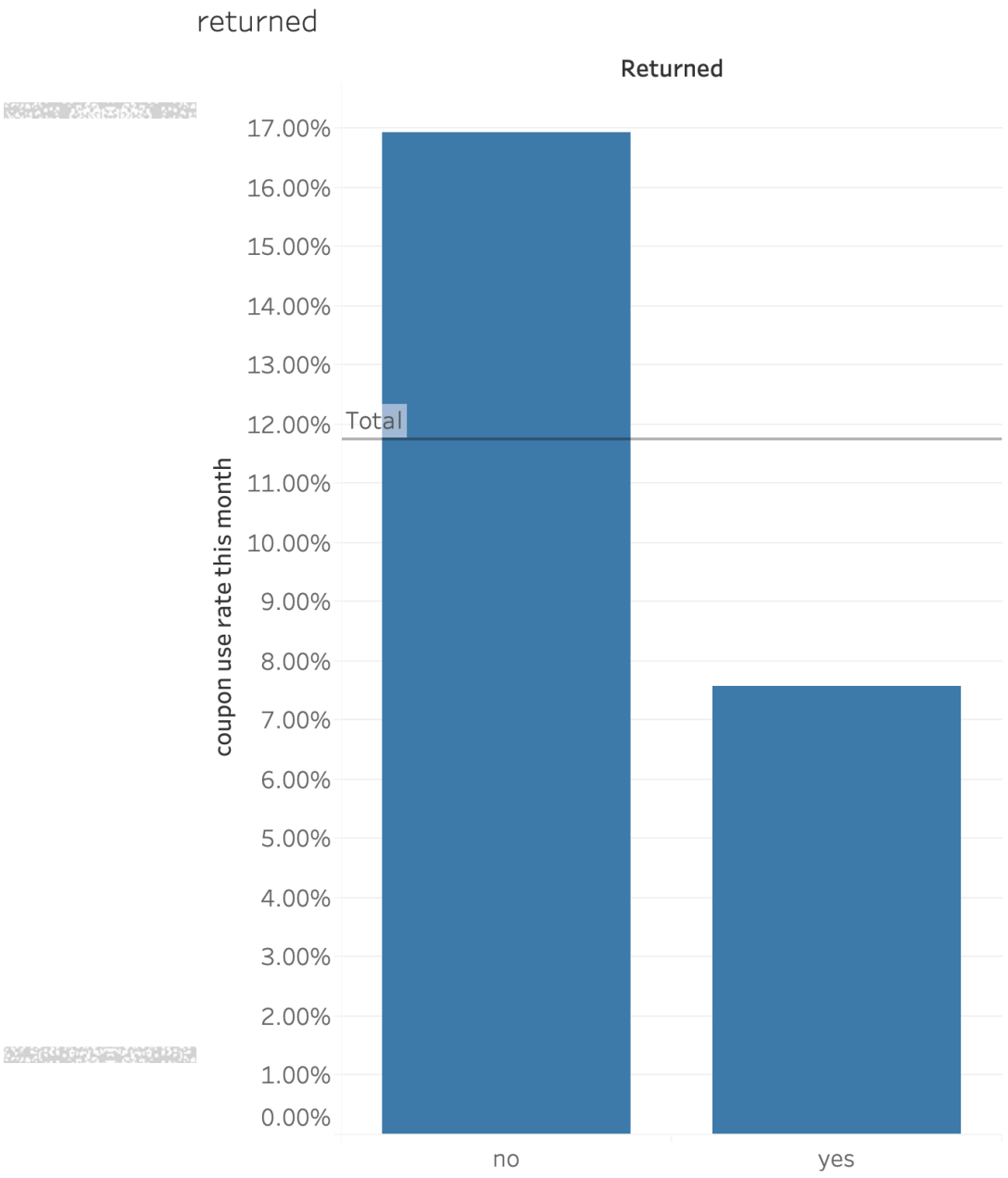
maritial



WHAT PEOPLE ARE MOST LIKELY TO USE NEW COUPON

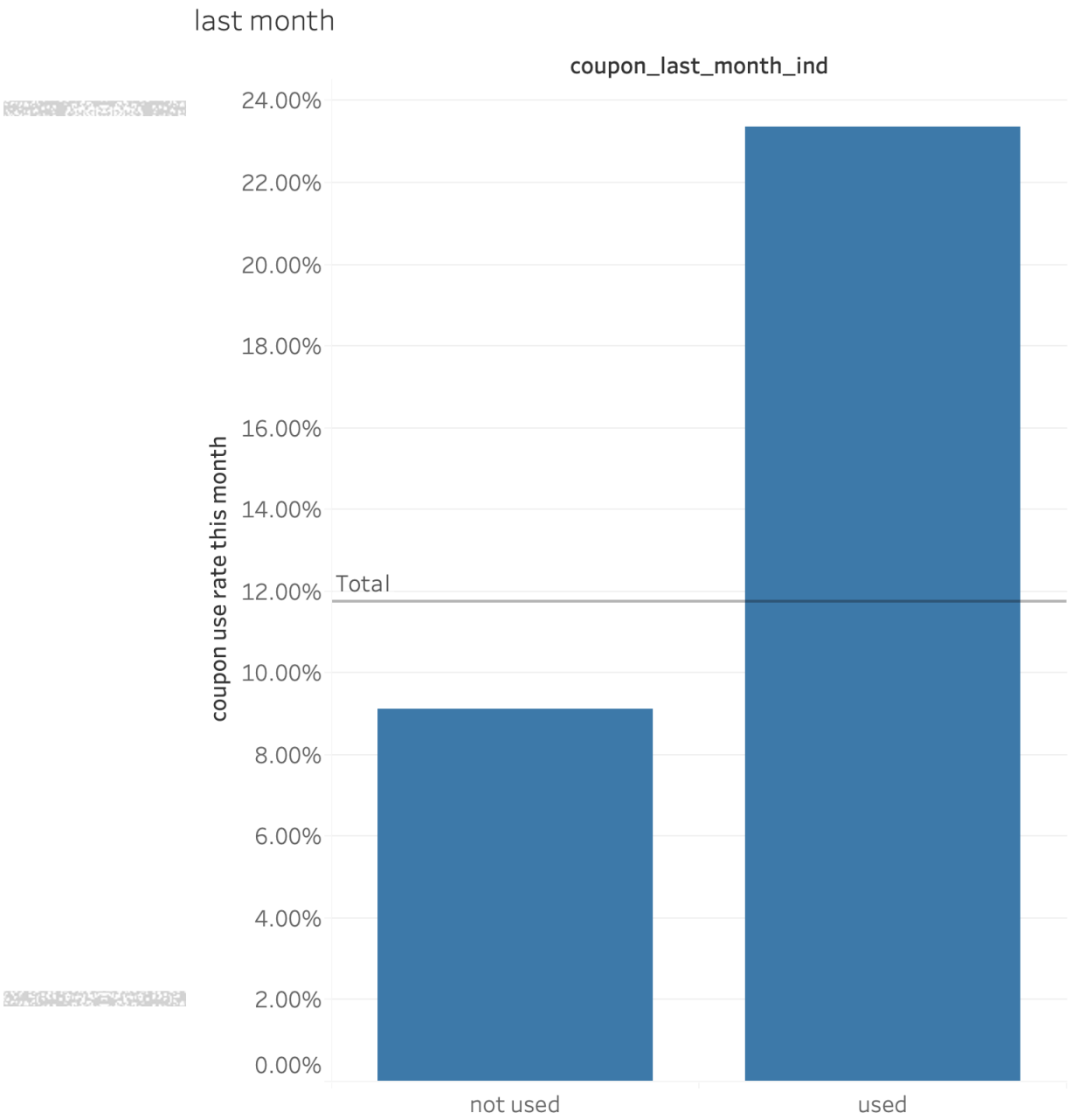
Among all marital status, single people is most likely to use coupon (+22%).

Married people is least likely to use coupon.



WHAT PEOPLE ARE MOST LIKELY TO USE NEW COUPON

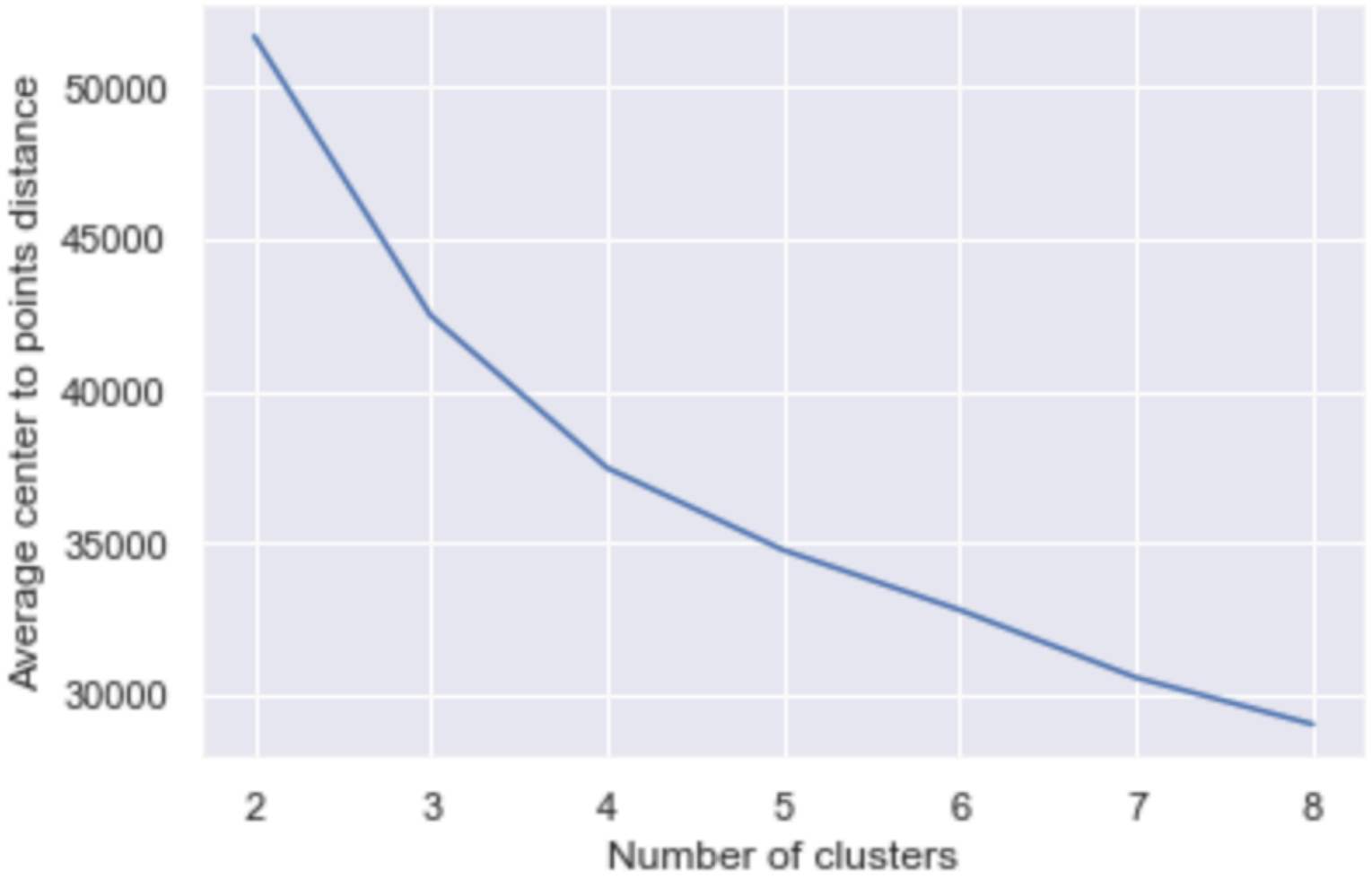
People who have not returned any items is more likely to use coupon (+41%).



WHAT PEOPLE ARE MOST LIKELY TO USE NEW COUPON

People who used coupon last month is more likely to use new coupon (+92%).

CLUSTERING CUSTOMERS: K-PROTOTYPE (K-MEANS)

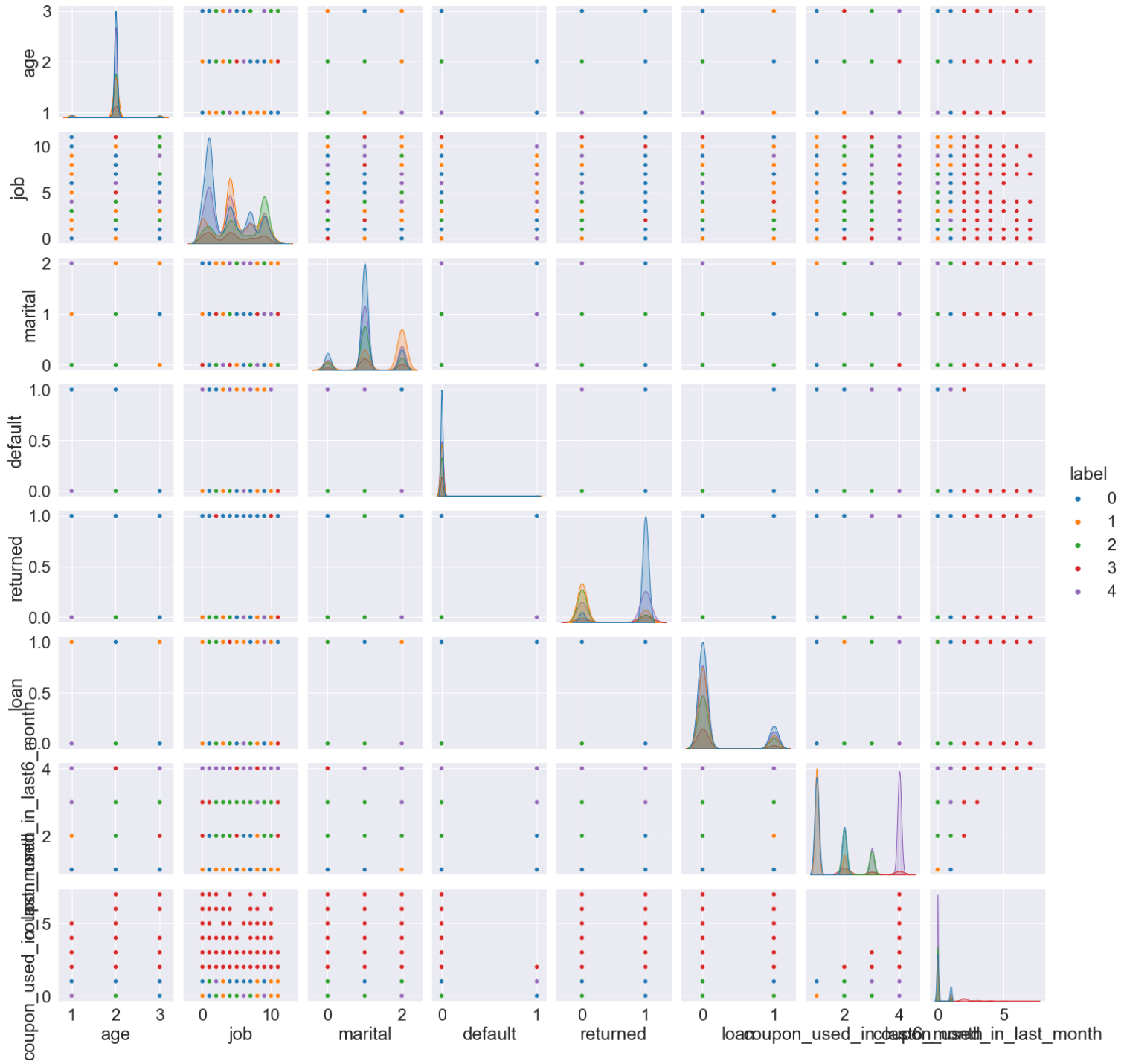


5 is the number of cluster when average point-centroid distance decreases linearly.

Also, centroid makes the most senses.

CLUSTERING SUMMARY

Cluster number	Percentage in total customers	Job	Marital	Returned before	Coupon used in last 6 months	Coupon used in last month
0	30.5%	blue collar	married	yes	1.33	0.24
1	21.8%	management	single	no	1.15	0.13
2	16.5%	technician	married	no	2.35	0.05
3	6.7%	management	married	yes	2.75	2.62
4	24.4%	blue collar	married	yes	3.80	0.03



CLUSTERING SUMMARY

- Last month, coupon is used very often among Cluster 3 (management, married, returned)
- Before last month, coupon is ed used very often among Cluster 4 (blue collar and technician, married)

PREDICT COUPON USAGE - METRIC

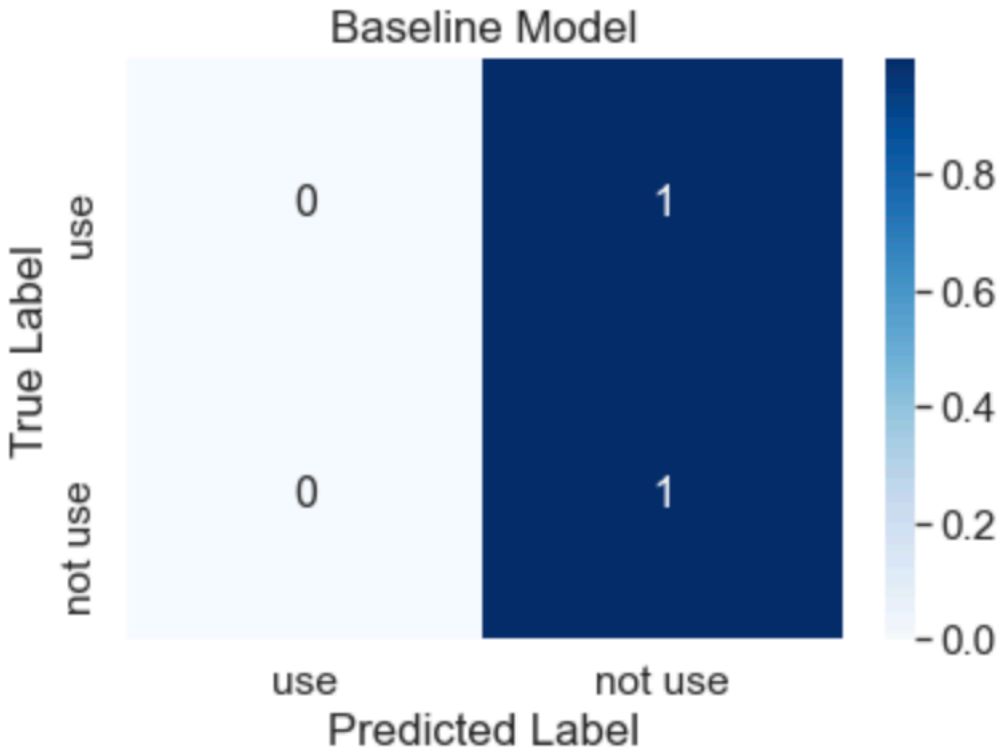
Assume there is no opportunity cost in giving out wrong coupon.

- Use Recall (minimize chance of miss-out potential customer)
- But this gives unbeatable baseline model (give everybody coupon)

Assume there is opportunity cost, assume cost the same as give right coupon

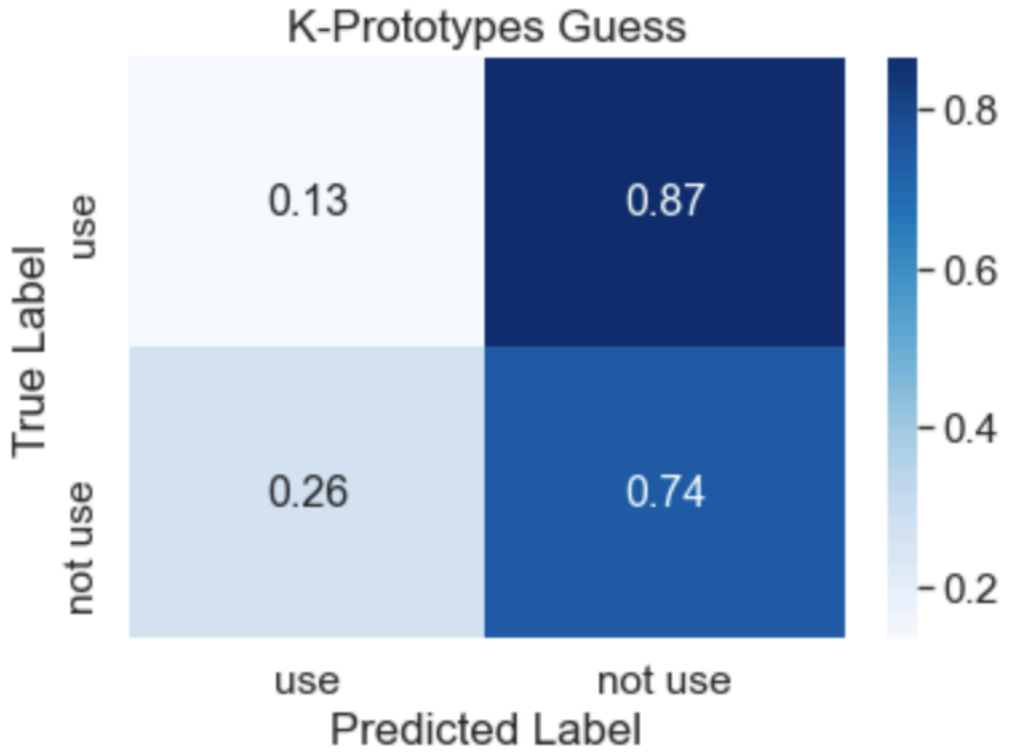
- Accuracy (treat + and – equally)

PREDICT COUPON USAGE - BASELINE



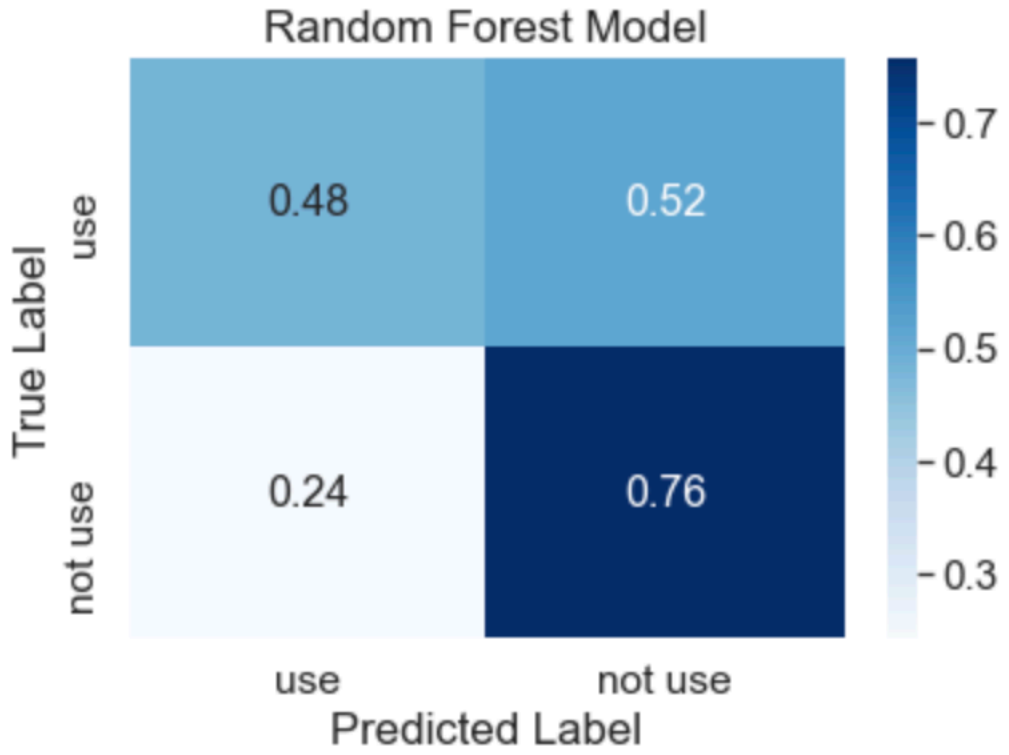
- We don't give anyone coupon.
- Accuracy is 0.882

PREDICT COUPON USAGE – K-PROTOTYPES



- Assume we don't know result from coupon usage
- Only give Cluster 4 coupon
- Accuracy is 0.670

PREDICT COUPON USAGE – RANDOM FOREST/LOGISTIC REGRESSION/BOOSTING



Logistic	RF	AdaBoost
0.663	0.724	0.674

- Assume we know result from coupon usage
- If for interpretation purpose I would use logistic regression, if for prediction purpose I would use tree-based method
- Best accuracy is 0.883

SUMMARY OF FINDINGS AND RECOMMENDATIONS

- New coupon is used less frequent than last month (-36%).
  - People returned items before are -66% less likely to use coupon.
  - Mid age group is -42% less likely to use coupon, where young and elder are +30% and +17% respectively.
- Overall, **non-working class** who has **never returned** before is more likely to use coupon this month.
- Re-evaluate if new coupon is biased on age and return policy
- Separate coupon strategy for different class based on clustering result. (i.e. management and technician/blue-collar)
- Accurate prediction is not achievable with current data source

APPENDIX 1

Correlation of age between age (26 and 60 inclusive) is not statistically significantly correlated to people who returned items before.

More rigorous proof can use Crammer V or Chi-square test.

https://datascience.stackexchange.com/questions/893/how-to-get-correlation-between-two-categorical-variable-and-a-categorical-variab

https://medium.com/@outside2SDs/an-overview-of-correlation-measures-between-categorical-and-continuous-variables-4c7f85610365

In [102]:
```
def age_func(age):
    if age >= 26 and age <= 60:
        return 1
    else:
        return 0
df_cleaned['returned'].apply(lambda x: 1 if x=='yes' else 0).corr(df_cleaned['age'].apply(age_func))
```

Out[102]:
```
0.1222442235237737
```

APPENDIX 2

Correlation of age between age (26 and 60 inclusive) is statistically significantly correlated to jobs, but not to people who returned items before.

More rigorous proof can use Crammer V or Chi-square test.

In [109]:
```
def job_func(x):
    if x=='retired' or x=='unemployed' or x=='student':
        return 1
    else:
        return 0
print(df_cleaned['job'].apply(job_func).corr(df_cleaned['age'].apply(age_func)))
print(df_cleaned['job'].apply(job_func).corr(df_cleaned['returned'].apply(lambda x: 1 if x=='yes' else 0)))
```

```
-0.3530316854578894
-0.1846706012576774
```