Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

Bank Marketing (Campaign)

08.20.2023

by Bizma Azeem and
Elif Nur Kemiksiz

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Executive Summary

- In this project, we conducted a data analysis on taxi usage in different cities and demographic groups in the United States to determine the most appropriate taxi investment for XYZ. The findings provide critical recommendations for the firm to consider when making an investment.

Problem Statement

- ABC Bank aims to launch a new term deposit product. However, their current marketing approach lacks precision, making it difficult to identify customers most likely to be interested. This leads to missed sales opportunities and potentially wasted marketing resources.

Approach

- By leveraging a targeted marketing approach, we aim to identify high-potential customers for the new term deposit product, optimizing outreach and maximizing sales success.

EDA

- The analysis revealed that certain customer segments and demographics are more likely to purchase term deposits. Additionally, some marketing channels are more effective in specific regions.

Summary Statistics of All Features:

```
Column Name: poutcome
poutcome
nonexistent    35563
failure         4252
success         1373
Name: count, dtype: int64

Column Name: education
education
university.degree      12168
high.school             9515
basic.9y                6045
professional.course     5243
basic.4y                4176
basic.6y                2292
unknown                 1731
illiterate                18
Name: count, dtype: int64

Column Name: y
y
no     36548
yes     4640
Name: count, dtype: int64
```

```
Column Name: default
default
no         32588
unknown     8597
yes            3
Name: count, dtype: int64

Column Name: housing
housing
yes        21576
no         18622
unknown      990
Name: count, dtype: int64

Column Name: loan
loan
no         33950
yes         6248
unknown      990
Name: count, dtype: int64

Column Name: contact
contact
cellular     26144
telephone    15044
Name: count, dtype: int64
```

```
Column Name: month
month
may    13769
jul     7174
aug     6178
jun     5318
nov     4101
apr     2632
oct      718
sep      570
mar      546
dec      182
Name: count, dtype: int64

Column Name: day_of_week
day_of_week
thu    8623
mon    8514
wed    8134
tue    8090
fri    7827
Name: count, dtype: int64
```

```
Column Name: job
job
admin.           10422
blue-collar       9254
technician        6743
services          3969
management        2924
retired           1720
entrepreneur      1456
self-employed     1421
housemaid         1060
unemployed        1014
student            875
unknown            330
Name: count, dtype: int64

Column Name: marital
marital
married     24928
single      11568
divorced     4612
unknown        80
Name: count, dtype: int64
```

Summary Statistics of All Features:

```
Column Name: age
count    41188.00000
mean        40.02406
std         10.42125
min         17.00000
25%         32.00000
50%         38.00000
75%         47.00000
max         98.00000
Name: age, dtype: float64
```

```
Column Name: cons.price.idx
count    41188.000000
mean        93.575664
std          0.578840
min         92.201000
25%         93.075000
50%         93.749000
75%         93.994000
max         94.767000
Name: cons.price.idx, dtype: float64
```

```
Column Name: duration
count    41188.000000
mean       258.285010
std        259.279249
min          0.000000
25%        102.000000
50%        180.000000
75%        319.000000
max       4918.000000
Name: duration, dtype: float64
```

```
Column Name: emp.var.rate
count    41188.000000
mean         0.081886
std          1.570960
min         -3.400000
25%         -1.800000
50%          1.100000
75%          1.400000
max          1.400000
Name: emp.var.rate, dtype: float64
```

```
Column Name: cons.conf.idx
count    41188.000000
mean       -40.502600
std          4.628198
min        -50.800000
25%        -42.700000
50%        -41.800000
75%        -36.400000
max        -26.900000
Name: cons.conf.idx, dtype: float64
```

```
Column Name: previous
count    41188.000000
mean         0.172963
std          0.494901
min          0.000000
25%          0.000000
50%          0.000000
75%          0.000000
max          7.000000
Name: previous, dtype: float64
```

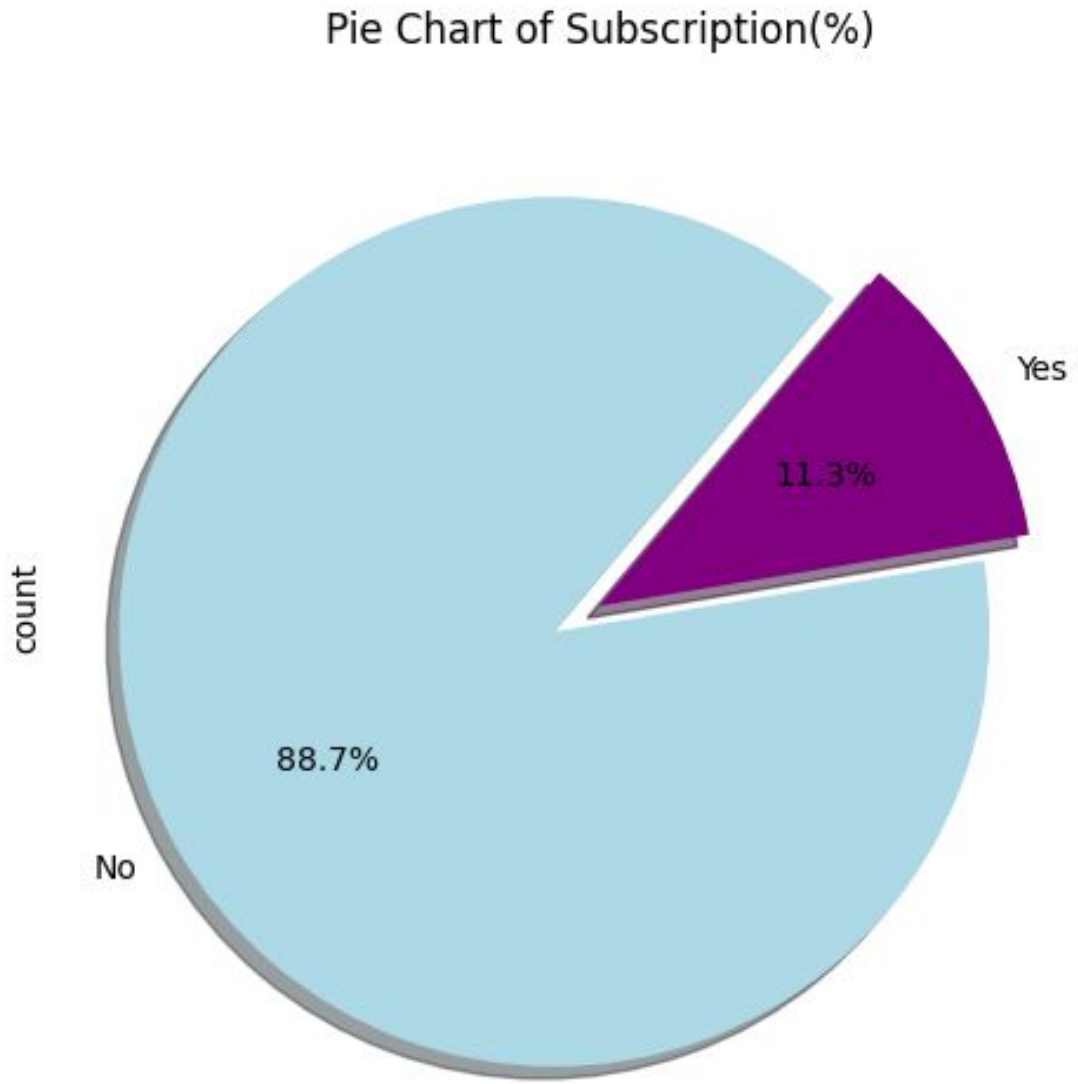
```
Column Name: nr.employed
count    41188.000000
mean      5167.035911
std         72.251528
min       4963.600000
25%       5099.100000
50%       5191.000000
75%       5228.100000
max       5228.100000
Name: nr.employed, dtype: float64
```

```
Column Name: euribor3m
count    41188.000000
mean         3.621291
std          1.734447
min          0.634000
25%          1.344000
50%          4.857000
75%          4.961000
max          5.045000
Name: euribor3m, dtype: float64
```

```
Column Name: pdays
count    41188.000000
mean       962.475454
std        186.910907
min          0.000000
25%        999.000000
50%        999.000000
75%        999.000000
max        999.000000
Name: pdays, dtype: float64
```

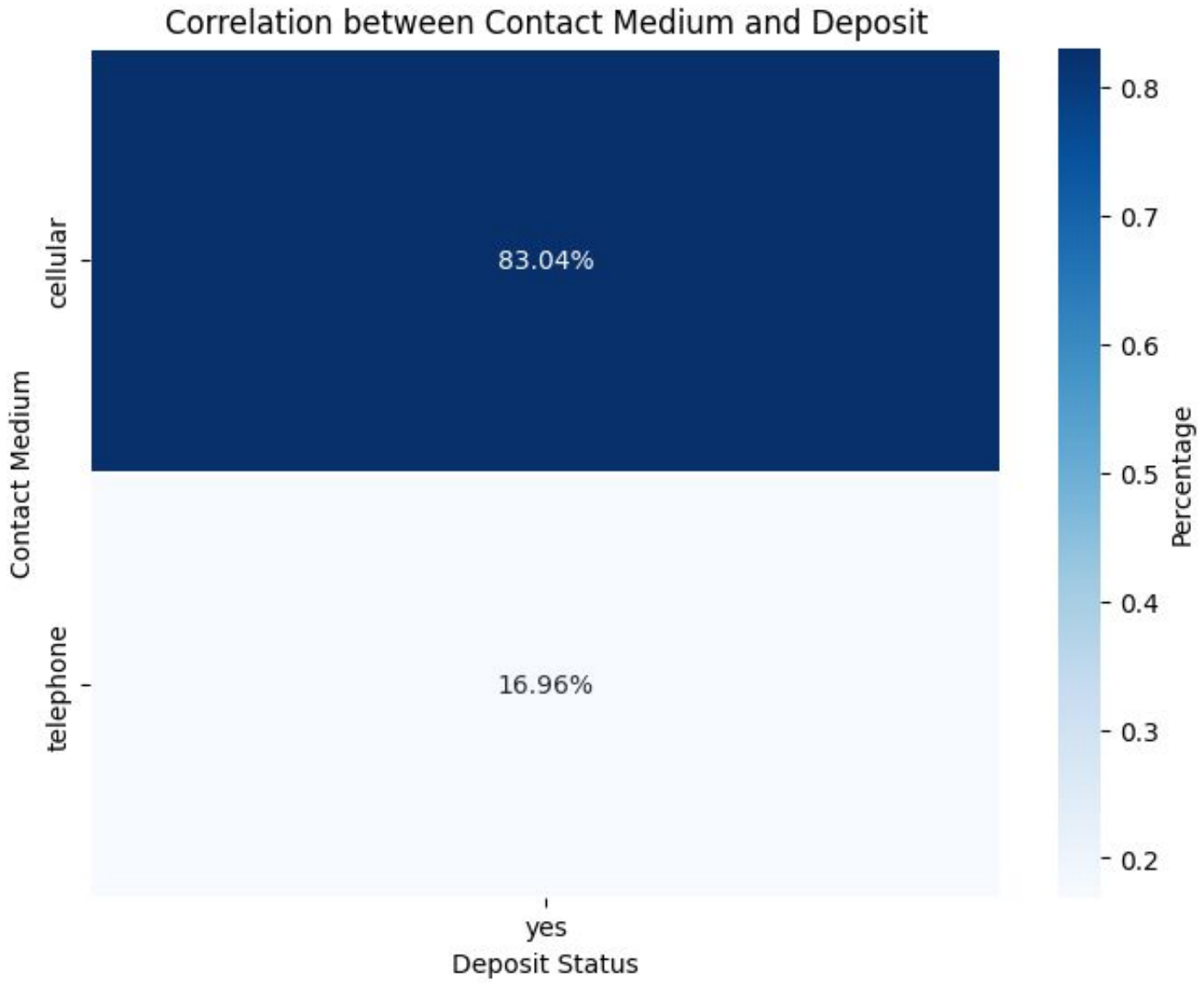
```
Column Name: campaign
count    41188.000000
mean         2.567593
std          2.770014
min          1.000000
25%          1.000000
50%          2.000000
75%          3.000000
max         56.000000
Name: campaign, dtype: float64
```

Pie Chart of Subscription rate:



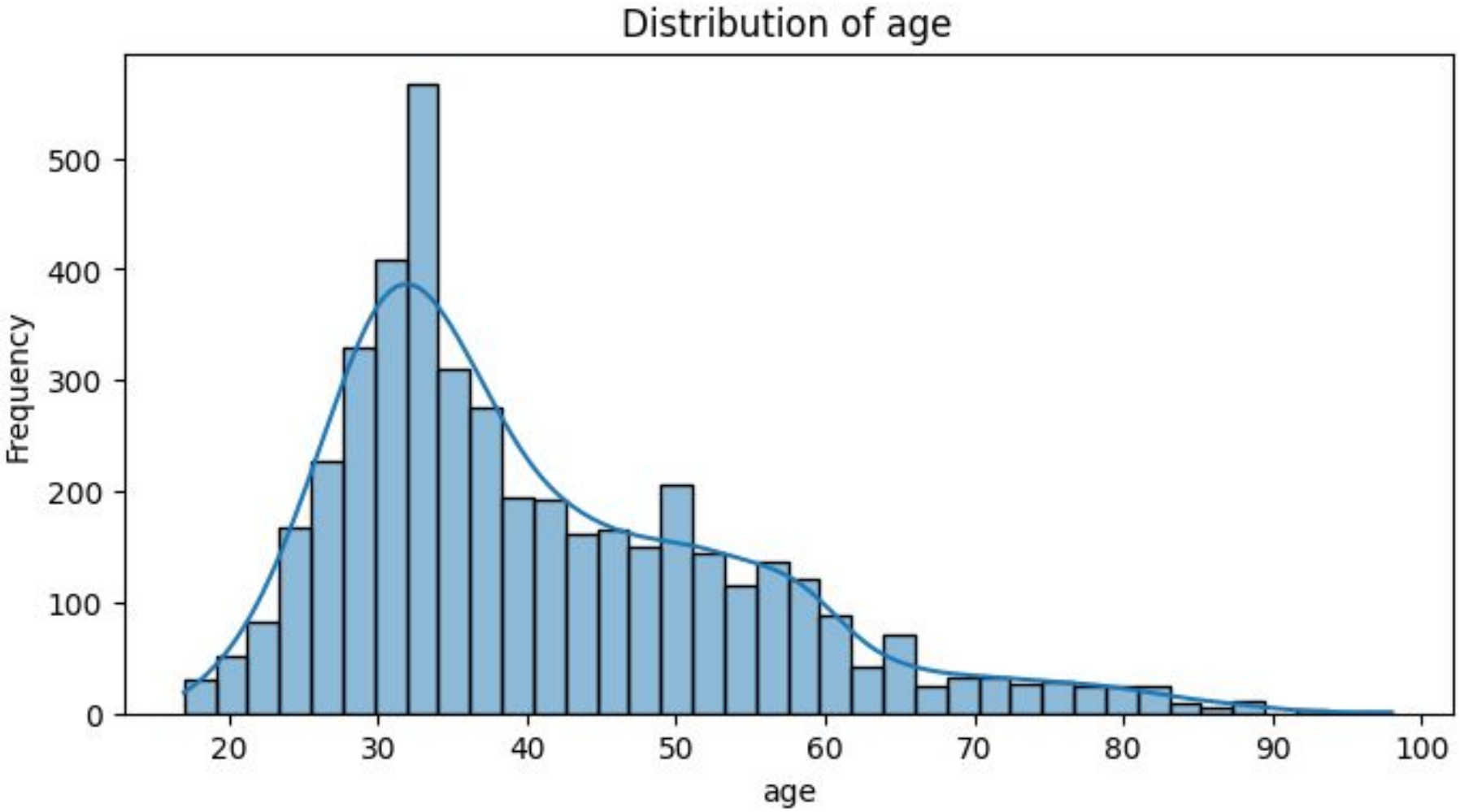
Data Glacier

Heatmap of Contact medium of Subscribed Customers:

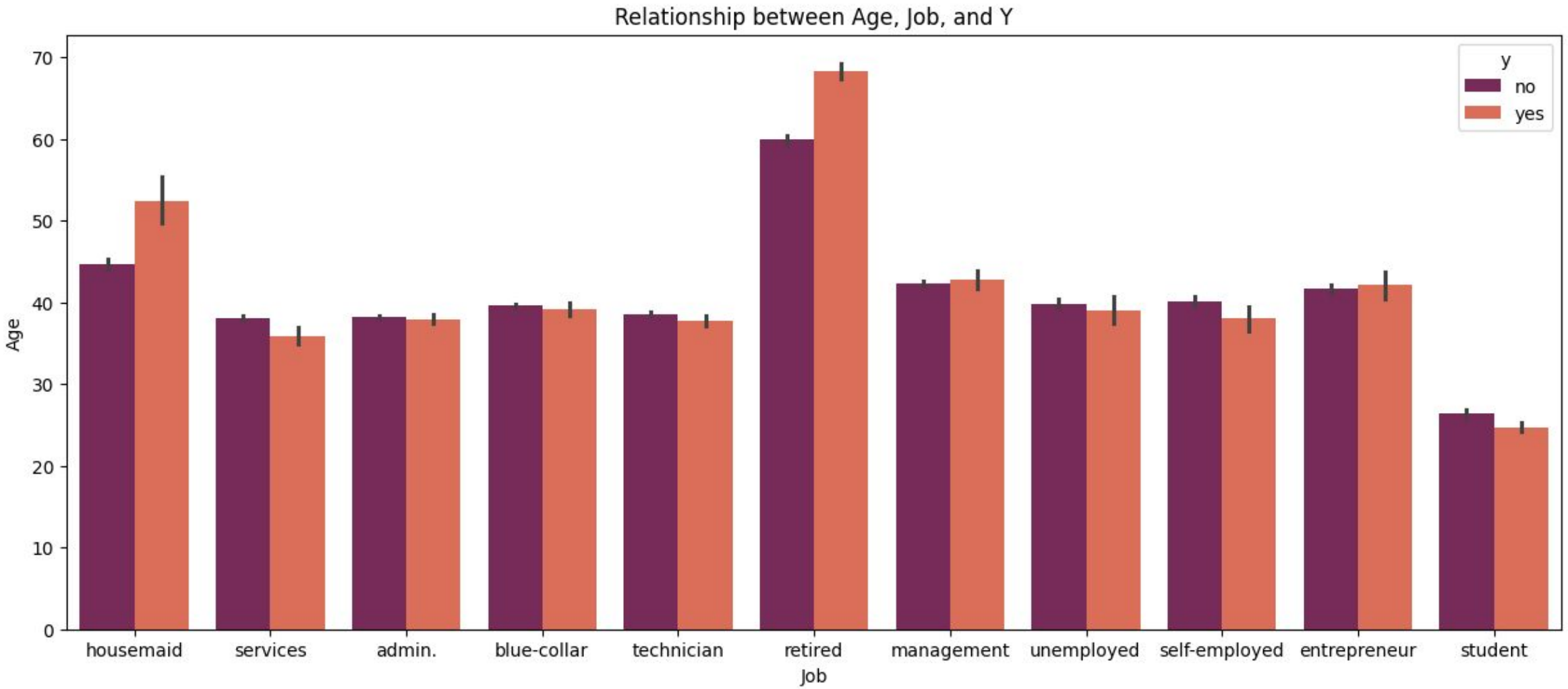


Data Glacier

Distribution of Age in Dataset:

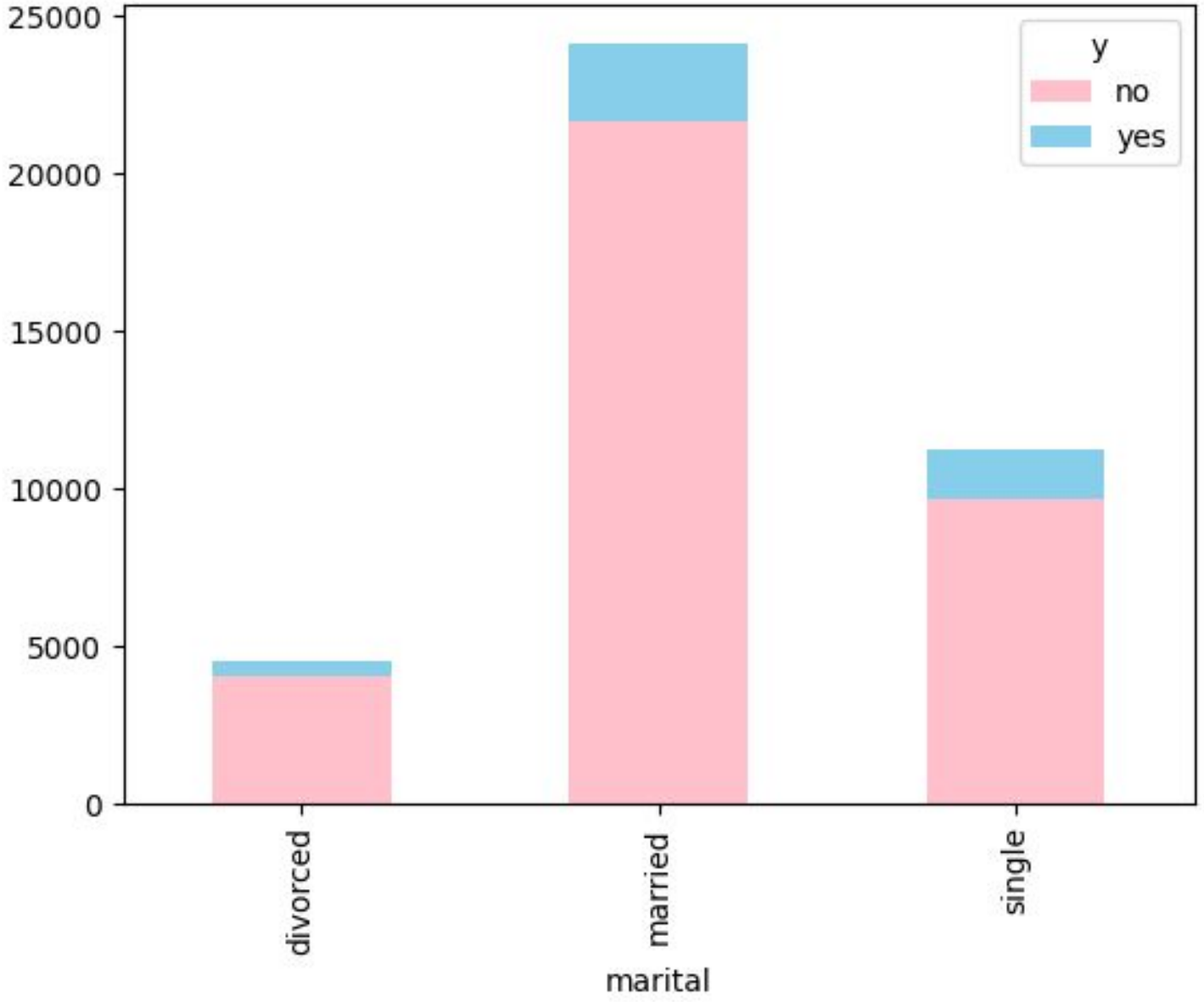


Relationship between Age, Job and Y:

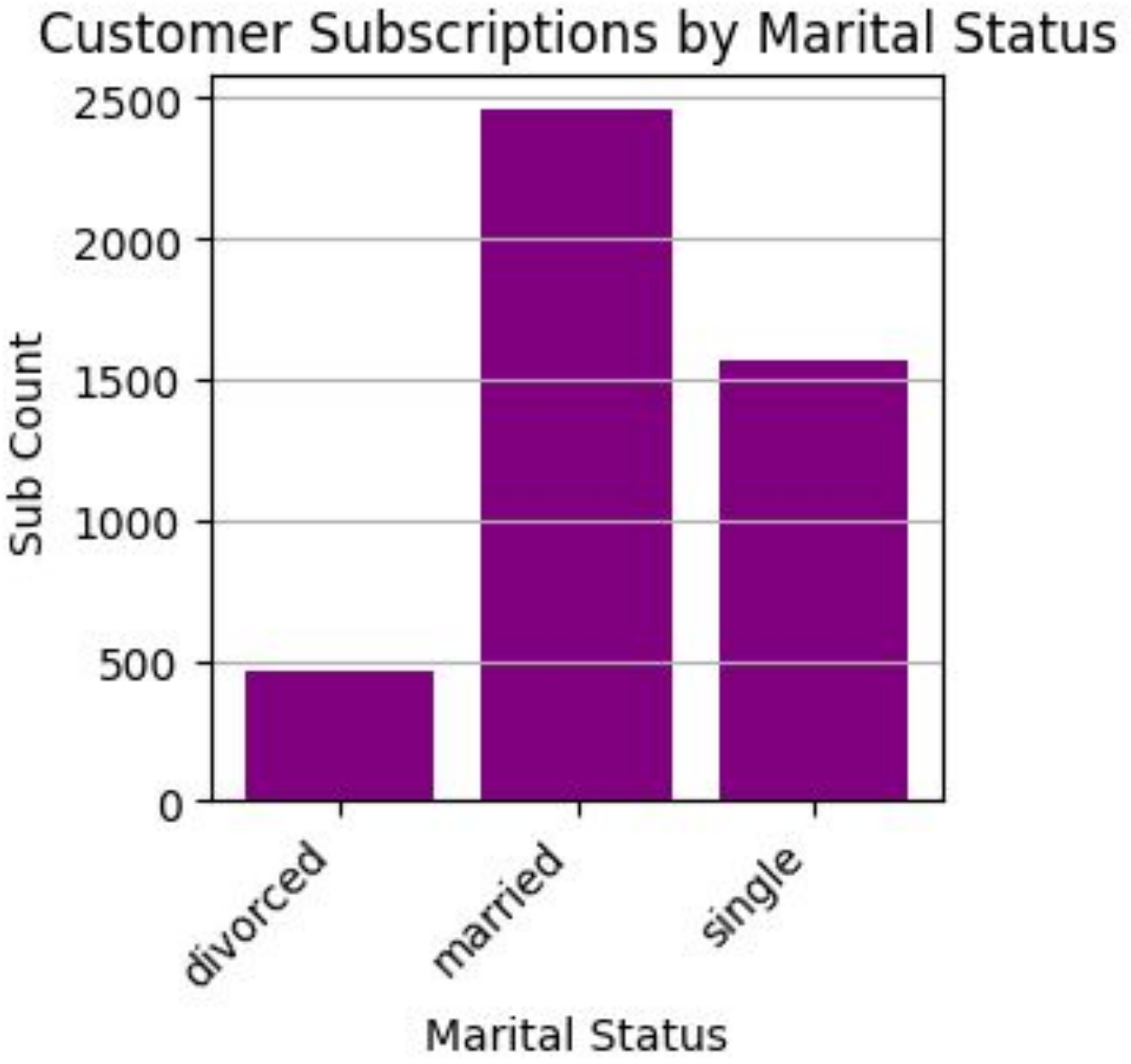


Data Glacier

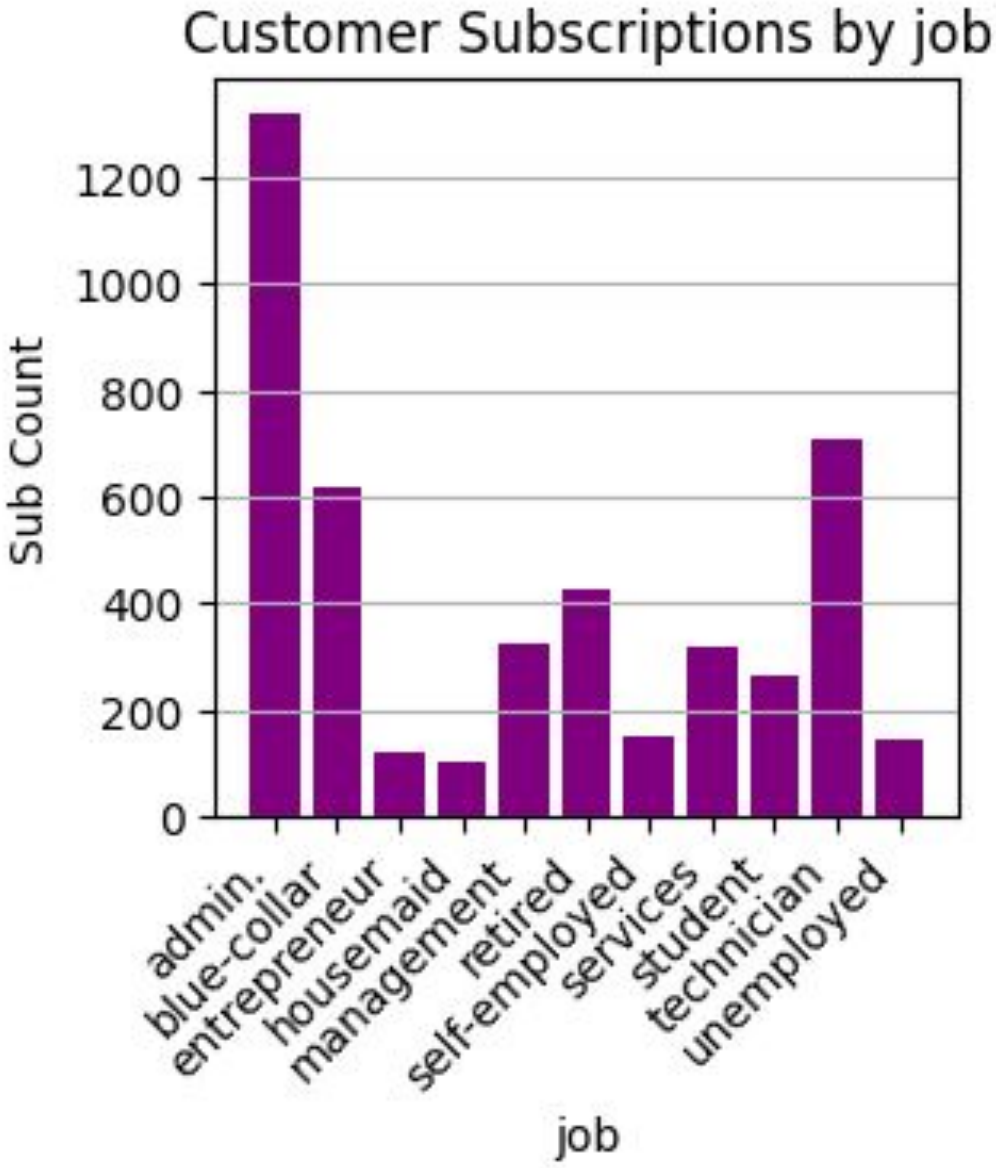
Distribution of marital status with respect to subscription status:



Distribution of marital status of Subscribed Customers:

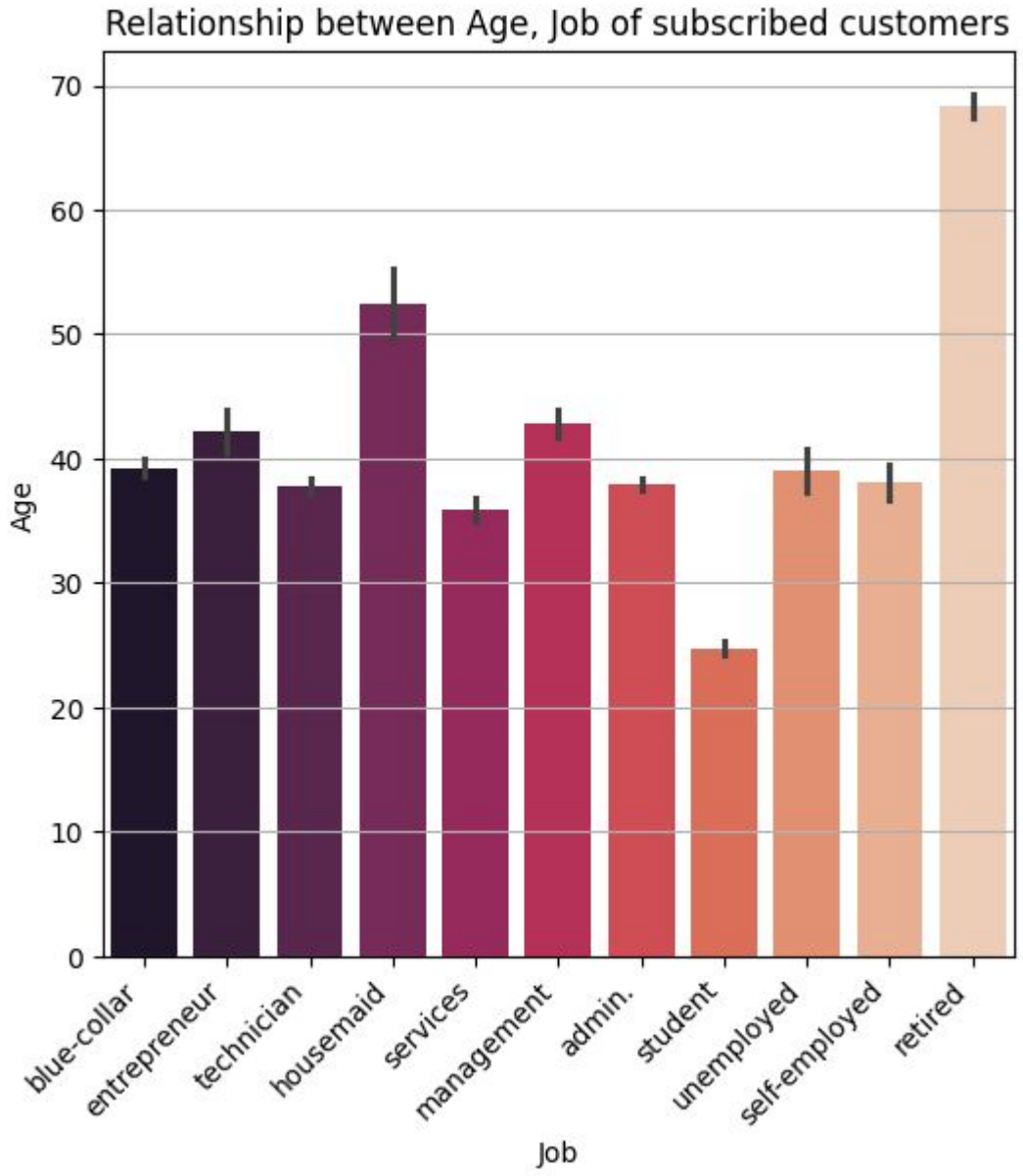


Distribution of Job of Subscribed Customers:



Distribution of day-of_week of Subscribed Customers:

Distribution of Job and Age of customers who subscribed:

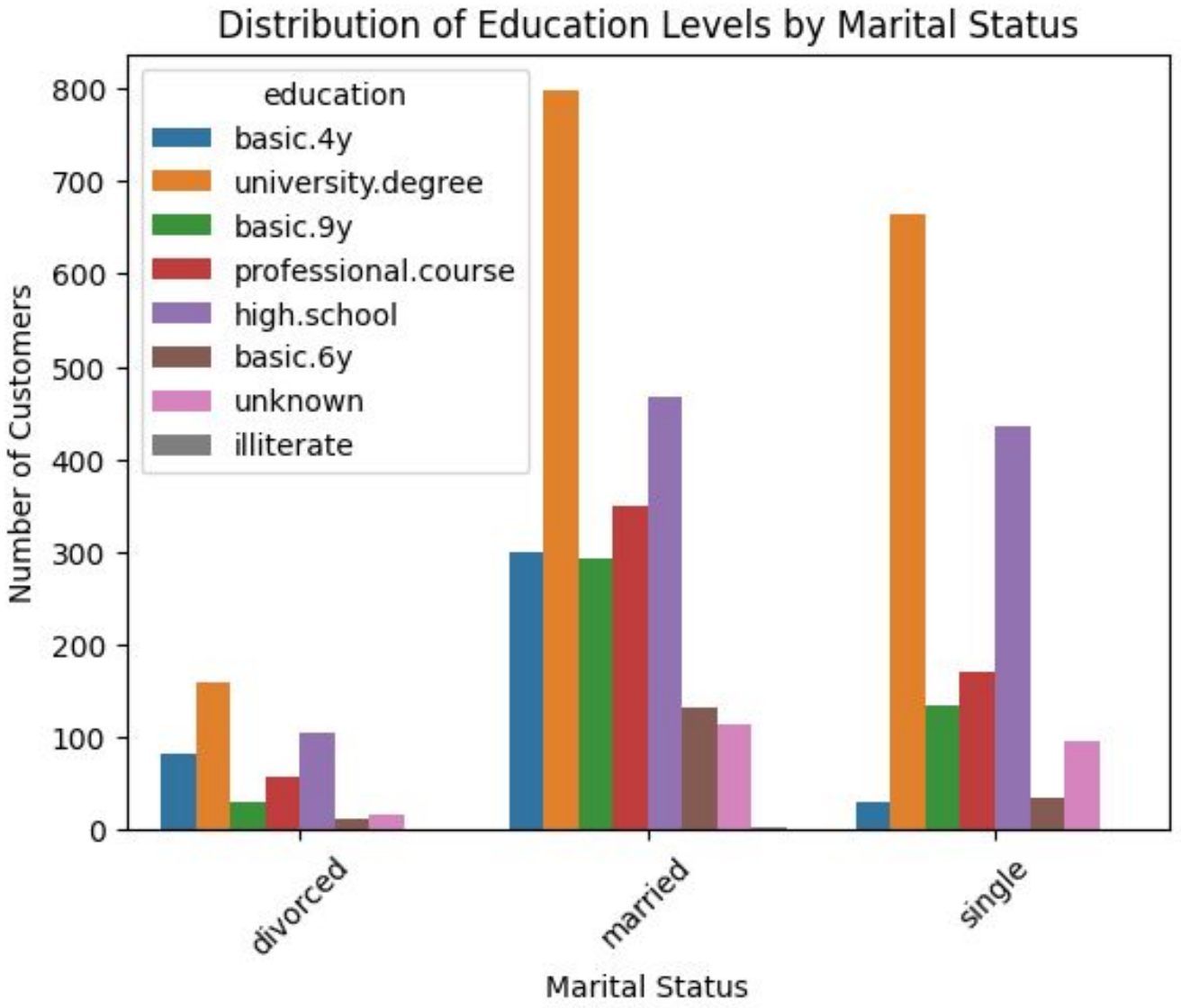


Hypothesis Testing for Targeted Marketing.

Hypothesis # 1 :
Do marital status and education level
influence customers' decisions to purchase term deposits?

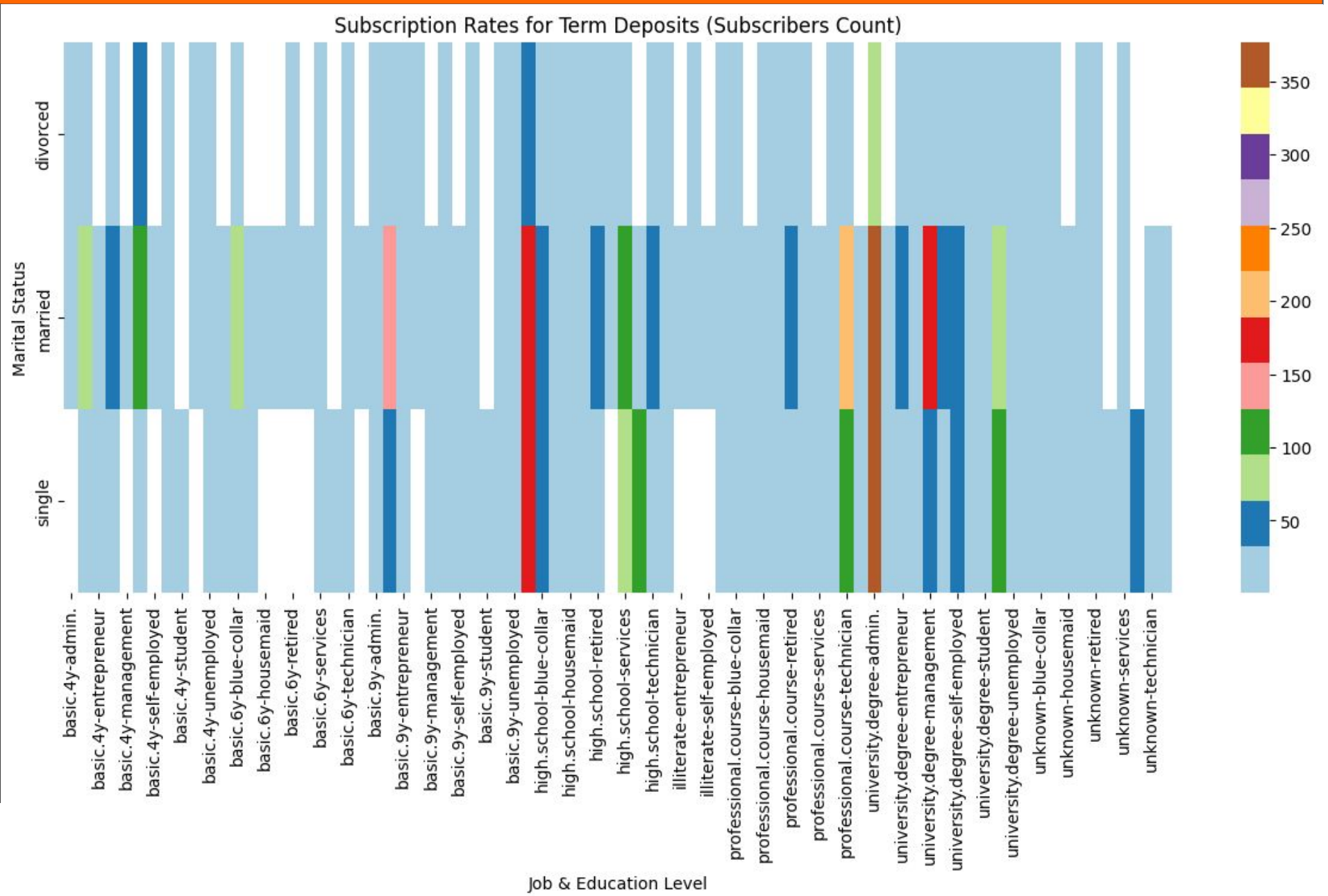
Distribution of Education level and Marital status of customers who subscribed:

Yes, There seem to have some relation between marital status and education level.
Single and married people with university level education are more likely
to make a purchase,
but this need further analysis.



Distribution of Education level and Marital status of customers who subscribed:

Factors like having a university degree, working in administration, and marital status (single or married) might be associated with a customer's decision to purchase a term deposit. Lets analyze it further



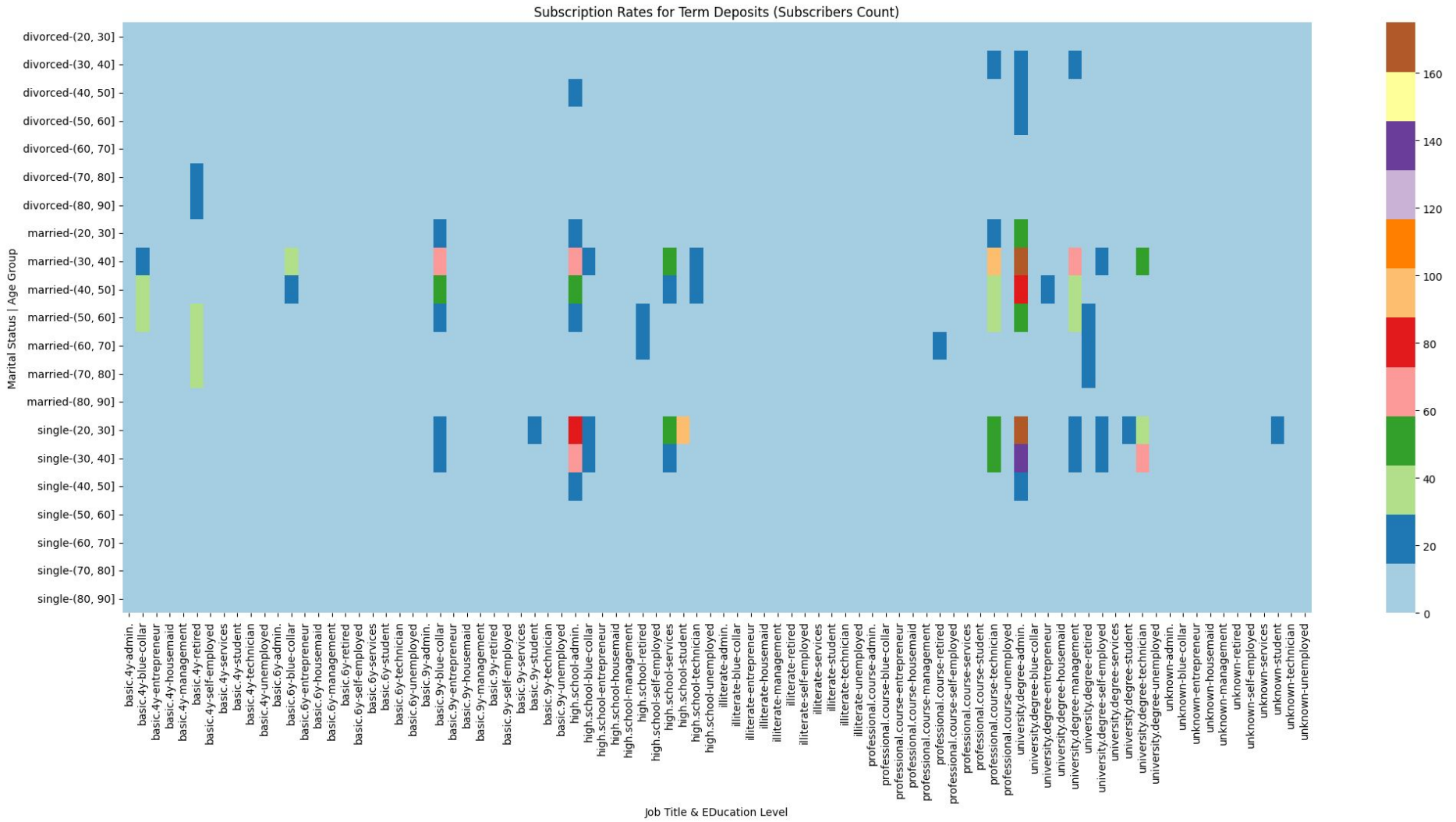
Data Glacier

Your Data Learning Partner

Distribution of Education level and Marital status of customers who subscribed:

The data suggests a trend where customers working in administration, who are single between the ages of 20 and 40 or married between 30 and 50, have higher subscription rates.

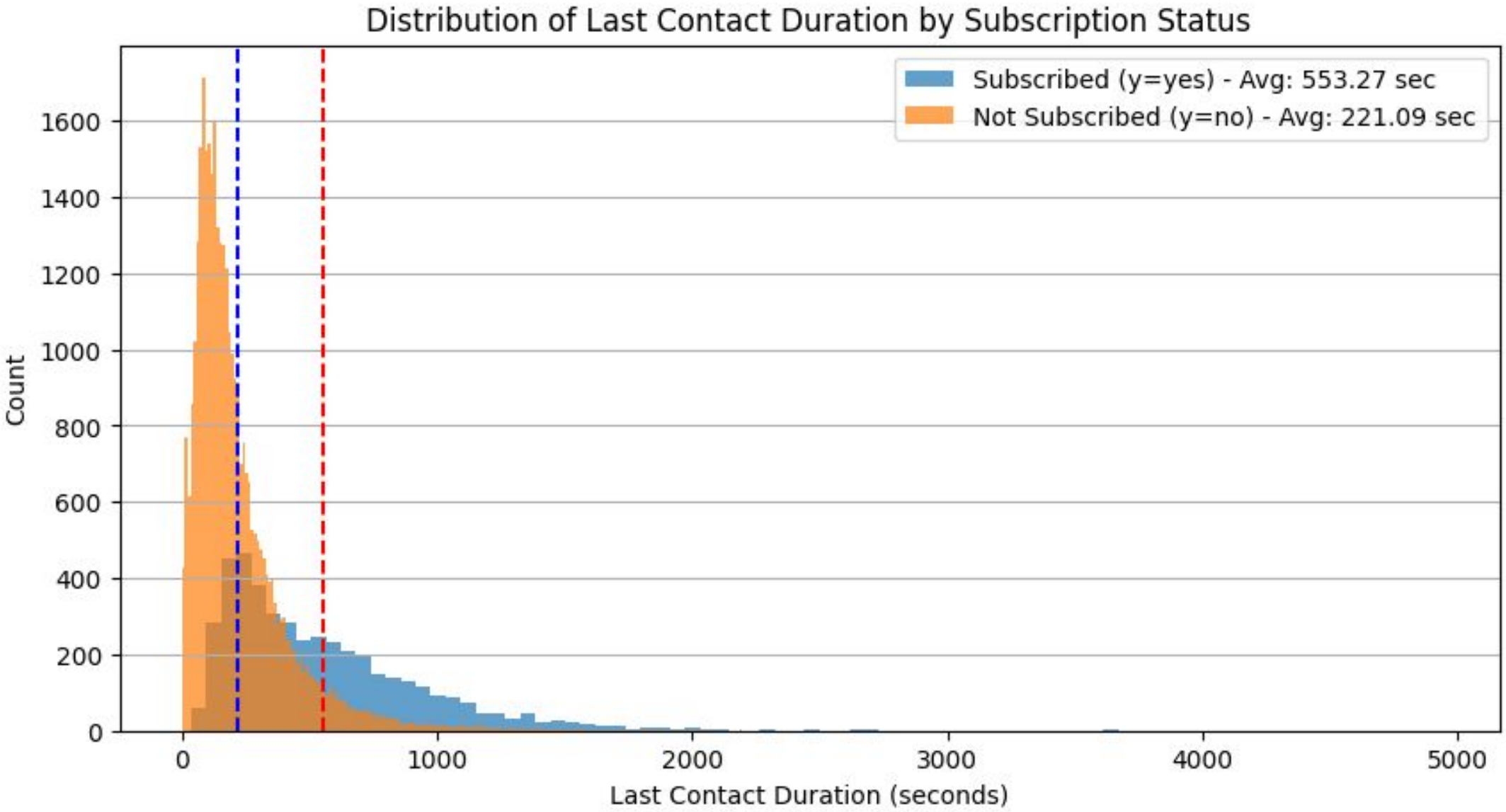
Hypothesis # 1 Accepted.



Data Glacier

Your Data Learning Partner

Distribution of call duration feature with respect to subscription status:



Data Glacier

Hypothesis # 2 :
Does contact medium and number of contacts during campaign influence customers' buying decision?

Distribution of Correlation between Contact Medium and Deposit status of customers who subscribed:

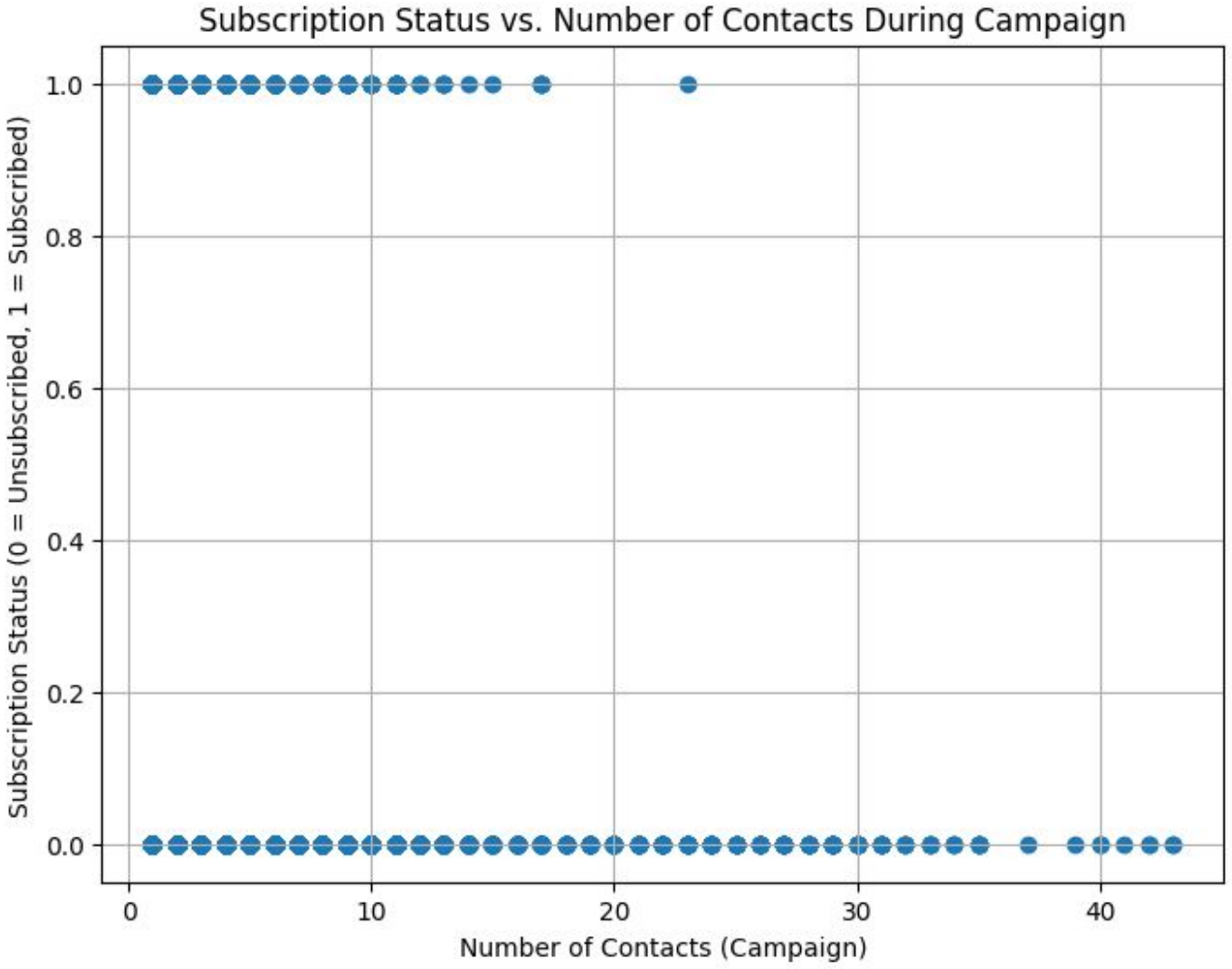
The data suggests a trend where the 'cellular' contact medium appears to be associated with a higher purchase rate compared to the 'telephone' medium.

Distribution of Correlation between Contact Medium and Deposit status of customers

While the current data doesn't show a strong correlation, a more comprehensive analysis could provide further insights into the impact of contact frequency.

While contact medium seem to influence but impact of contact frequency needs further analysis.

Hypothesis # 2 Rejected.

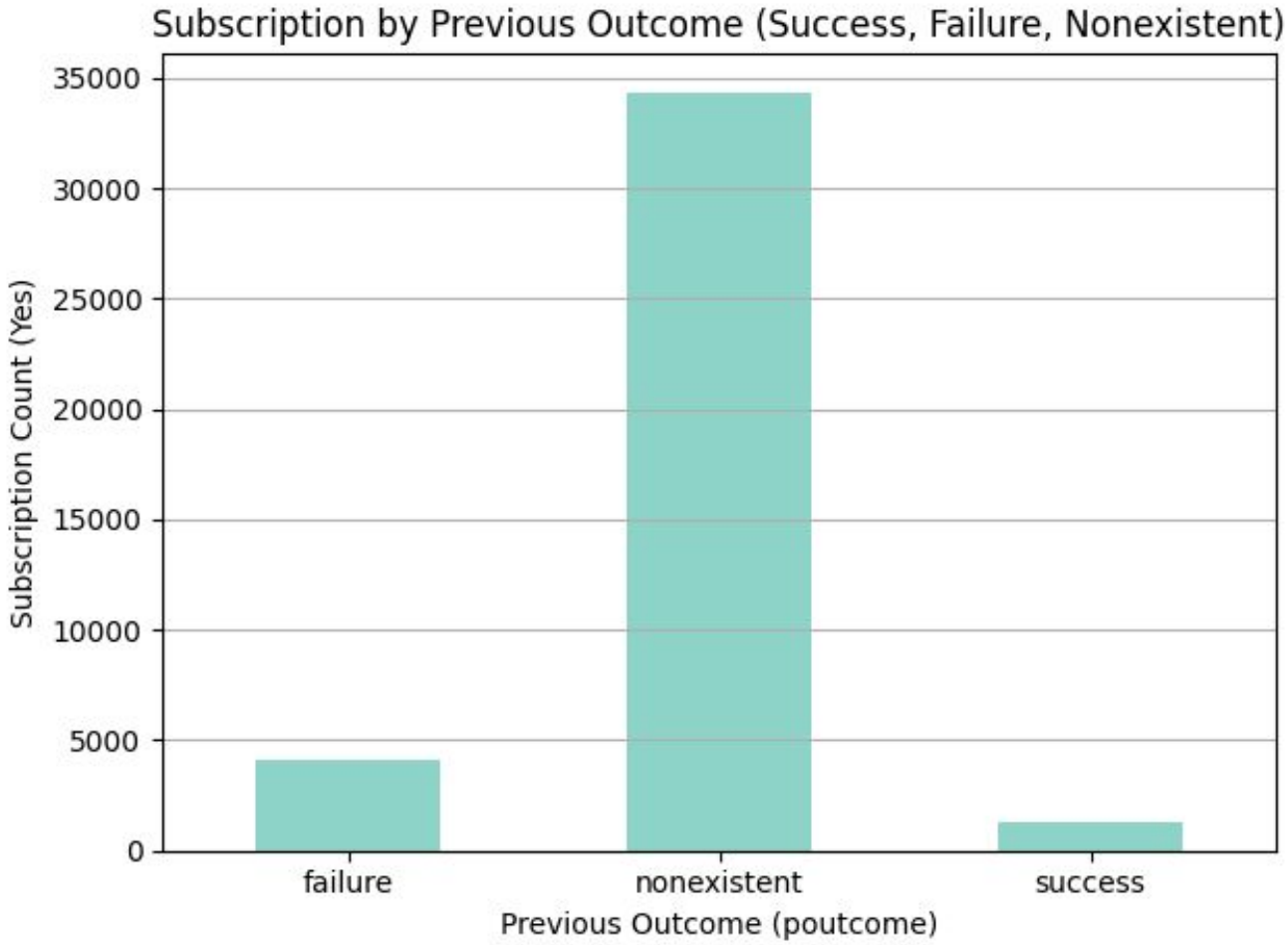


Hypothesis # 3 :

Customers who have been previously contacted (e.g., through phone calls, marketing campaigns) are more likely to subscribe to the service compared to new customers who have not been contacted before?

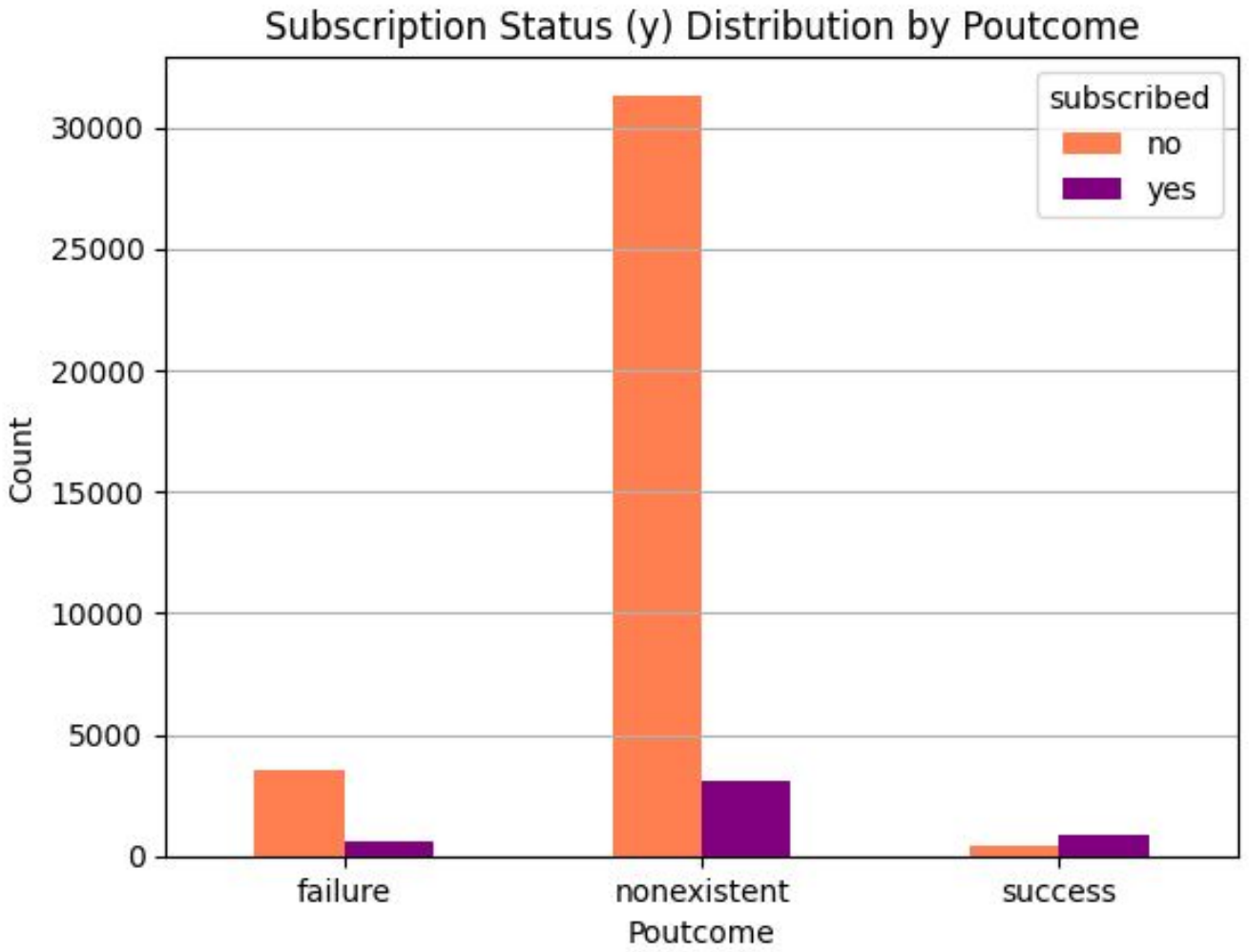
Subscription by Previous Outcome:

Previous outcome does not seem to influence subscription decision.Further investigation needed.



Subscription Status Distribution by Poutcome:

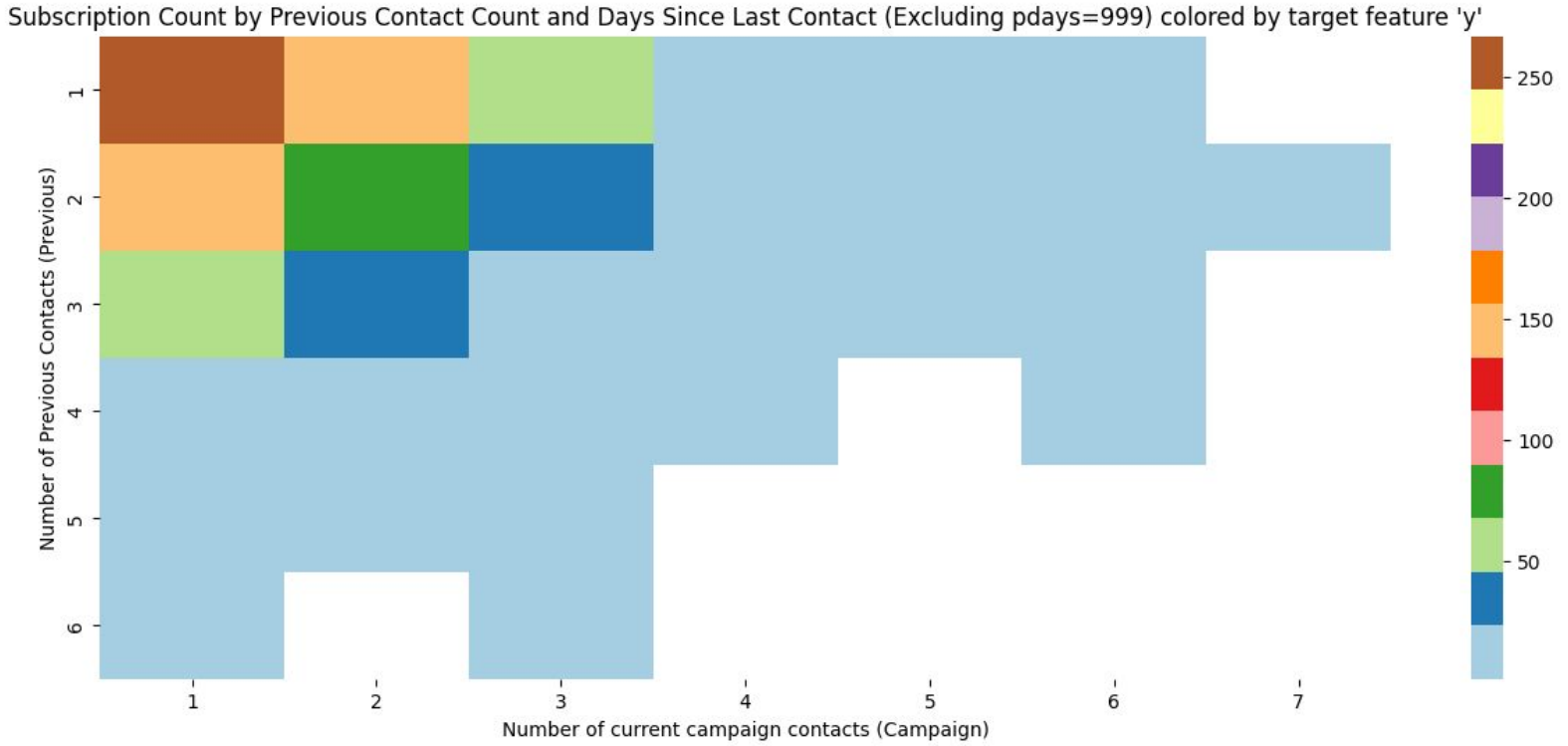
pdays also does not seem to influence subscription decision. Further investigation needed



Subscription Count by Previous Contact Count and Days Since Last Contact:

As the heatmap depicts, previous contacts does seem to influence buyers' decision.

Hypothesis # 3 Approved!



Data Glacier

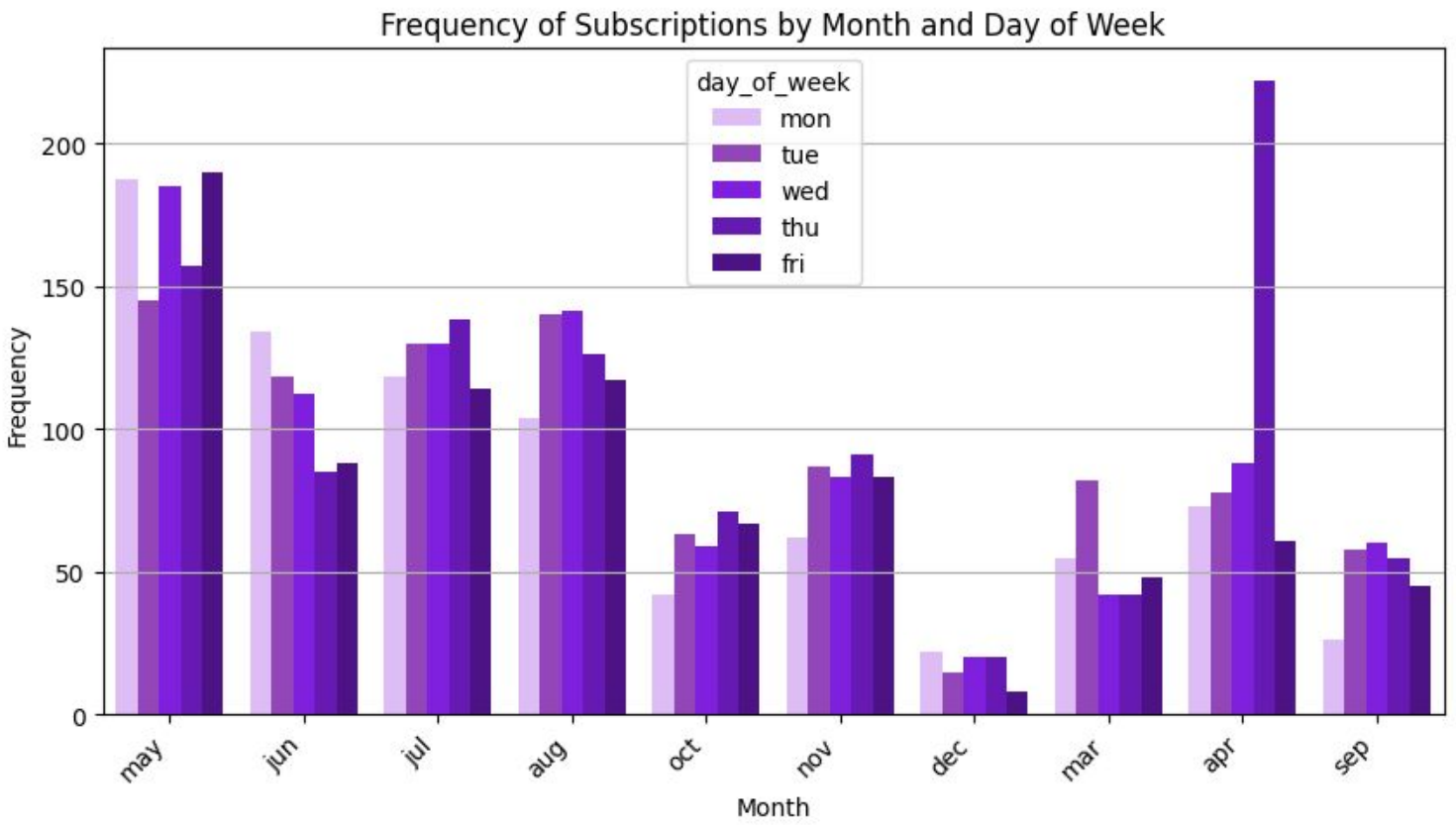
Your Data Learning Partner

Hypothesis # 4 :
Do weekdays within certain months lead to higher subscription rates?

Frequency of Subscriptions by Month and Day of Week:

There is a potential association between weekdays in May and Thursdays in April with higher subscription rates compared to other weekdays in different months.

Hypothesis # 4 Approved



Data Glacier
Your Data Learning Partner

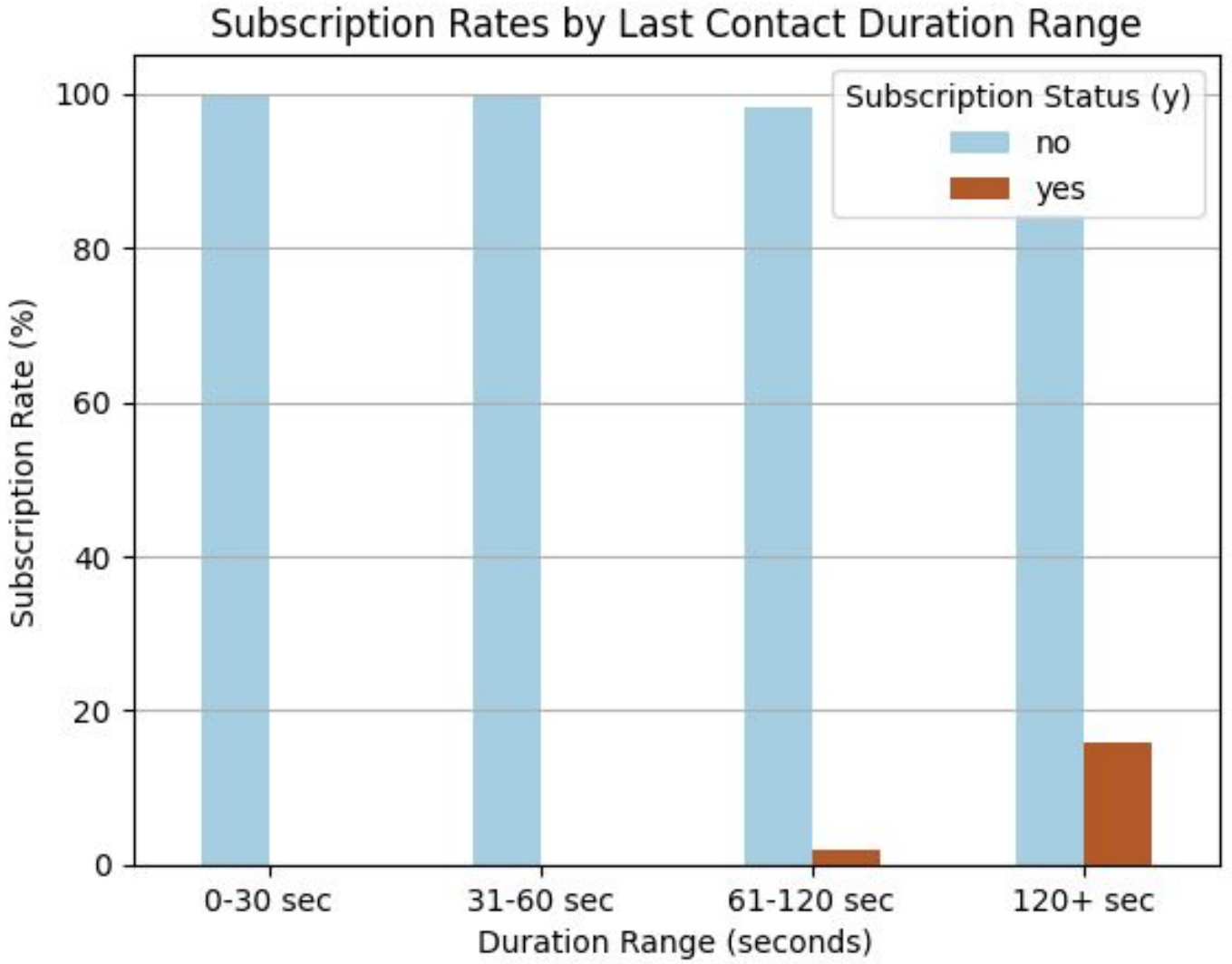
Hypothesis # 5 :
Does longer call duration ultimately influence buyers' buying decisions?

Subscription Status Distribution by Poutcome:

Duration is an important feature as it directly influence target variable 'y'. i.e If call duration is 0, then 'y' is also 0(No).

The graph shows that longer duration can ultimately influence buyers' decision.But further investigation is needed.

Hypothesis # 5 Accepted



Data Glacier

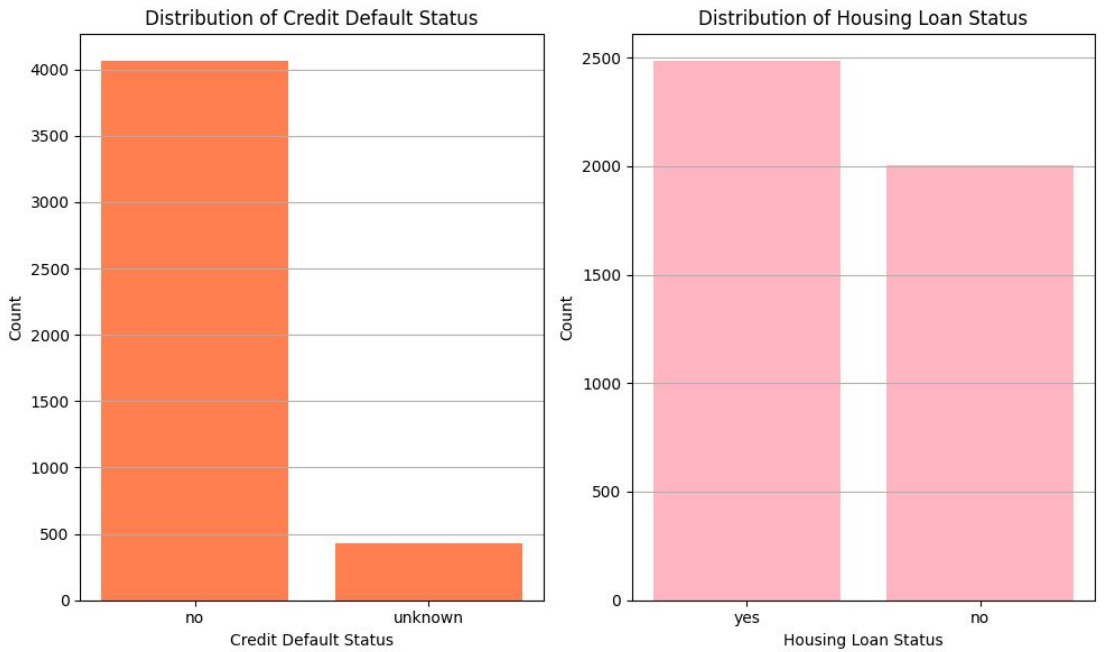
Hypothesis # 6 :

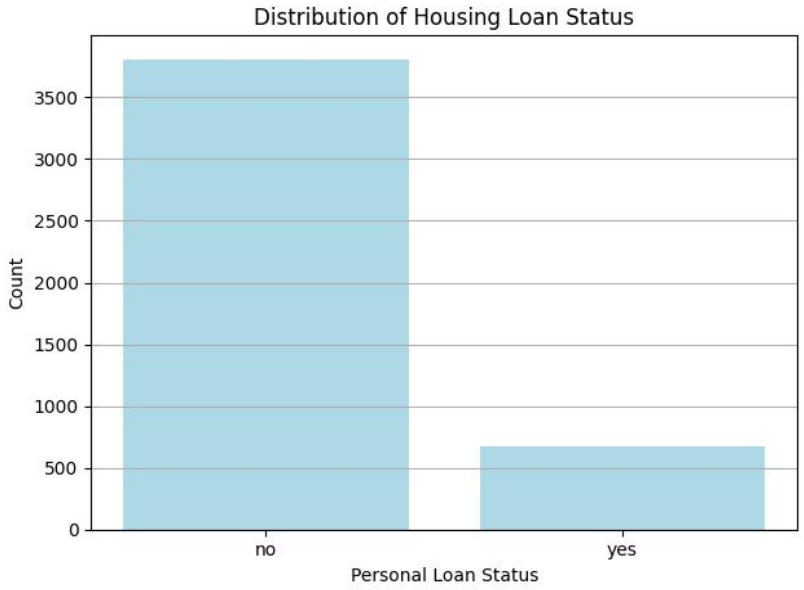
Customers with default status , personal and housing loans are less likely to subscribe.

Distribution of Credit Default Status, Distribution of Housing Loan Status, Distribution of Housing Loan Status

Although buyers with personal loan or default status were less likely to buy term deposit. but that's not the case for housing loan. Subscribed customers dataset have a high number of buyers who have housing loan. So, they can be a potential candidates for target marketing.

Hypothesis # 6 Rejected.



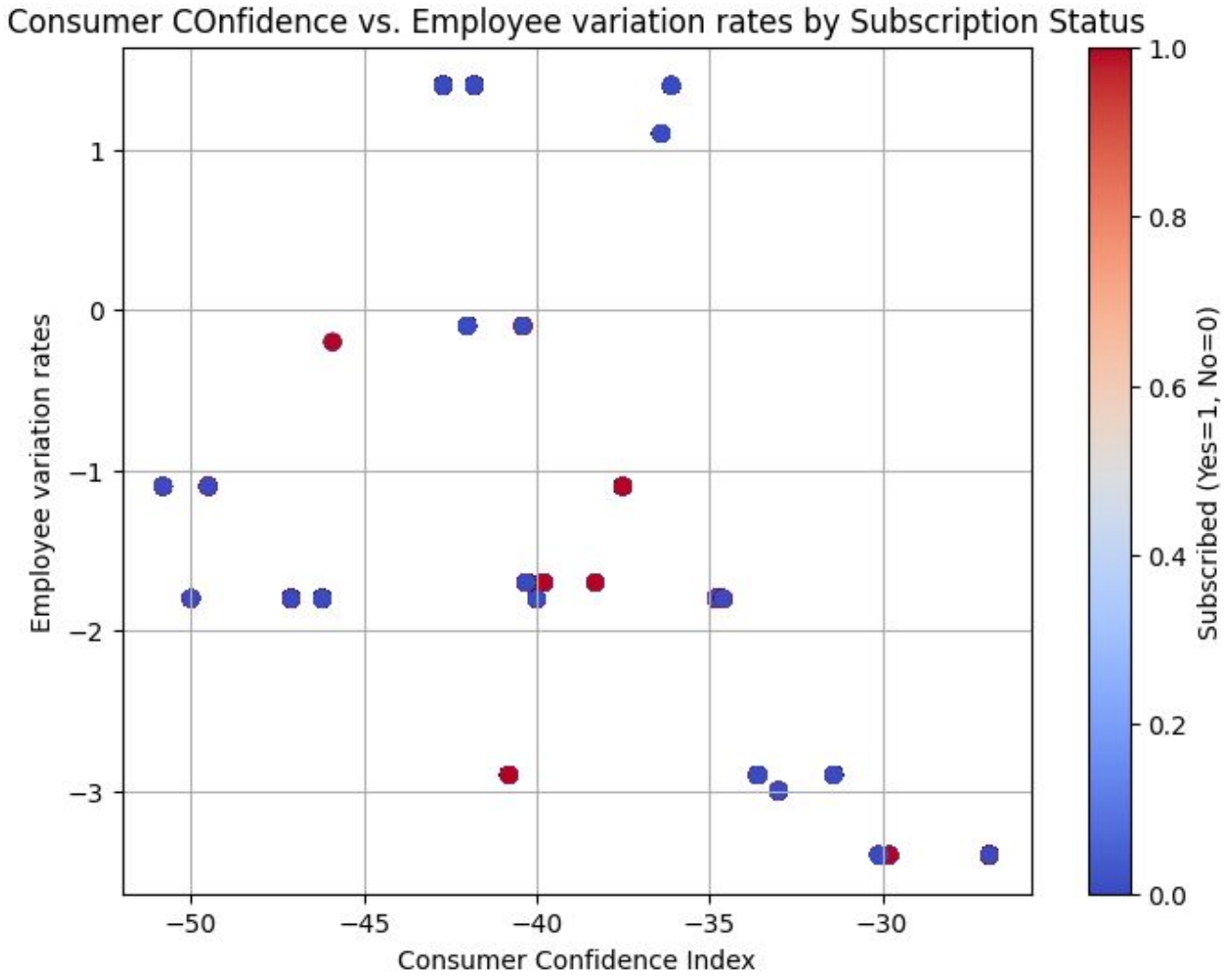


Data Glacier
Your Data Learning Partner

Hypothesis # 7 :
Consumers' economical factors affect their buying' decisions.

Consumer COnfidence vs. Employee variation rates by Subscription Status

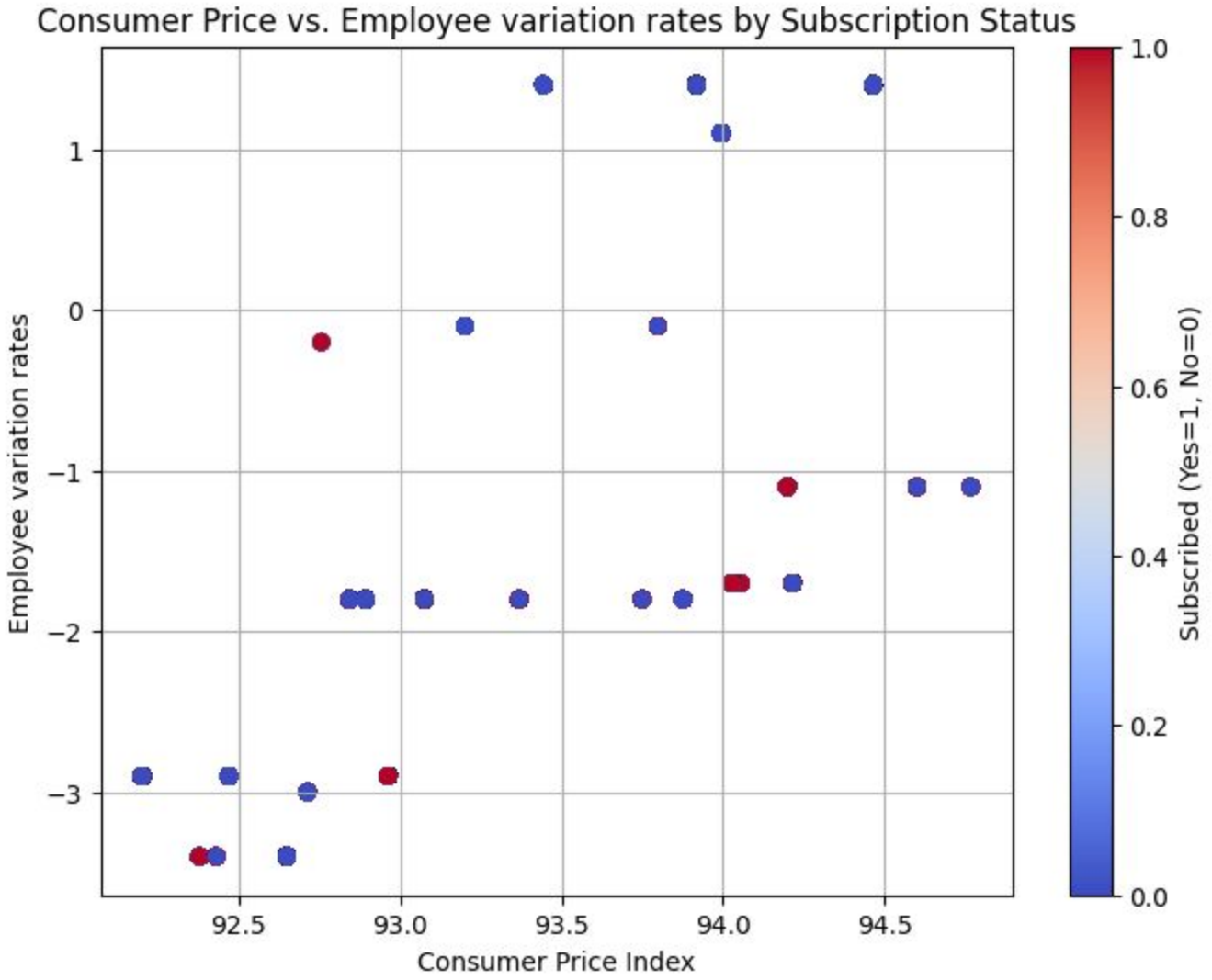
This graph depicts that customers with low emp.var.rate and high cons.conf.idx are more likely to purchase.



Data Glacier

Consumer Price vs. Employee variation rates by Subscription Status

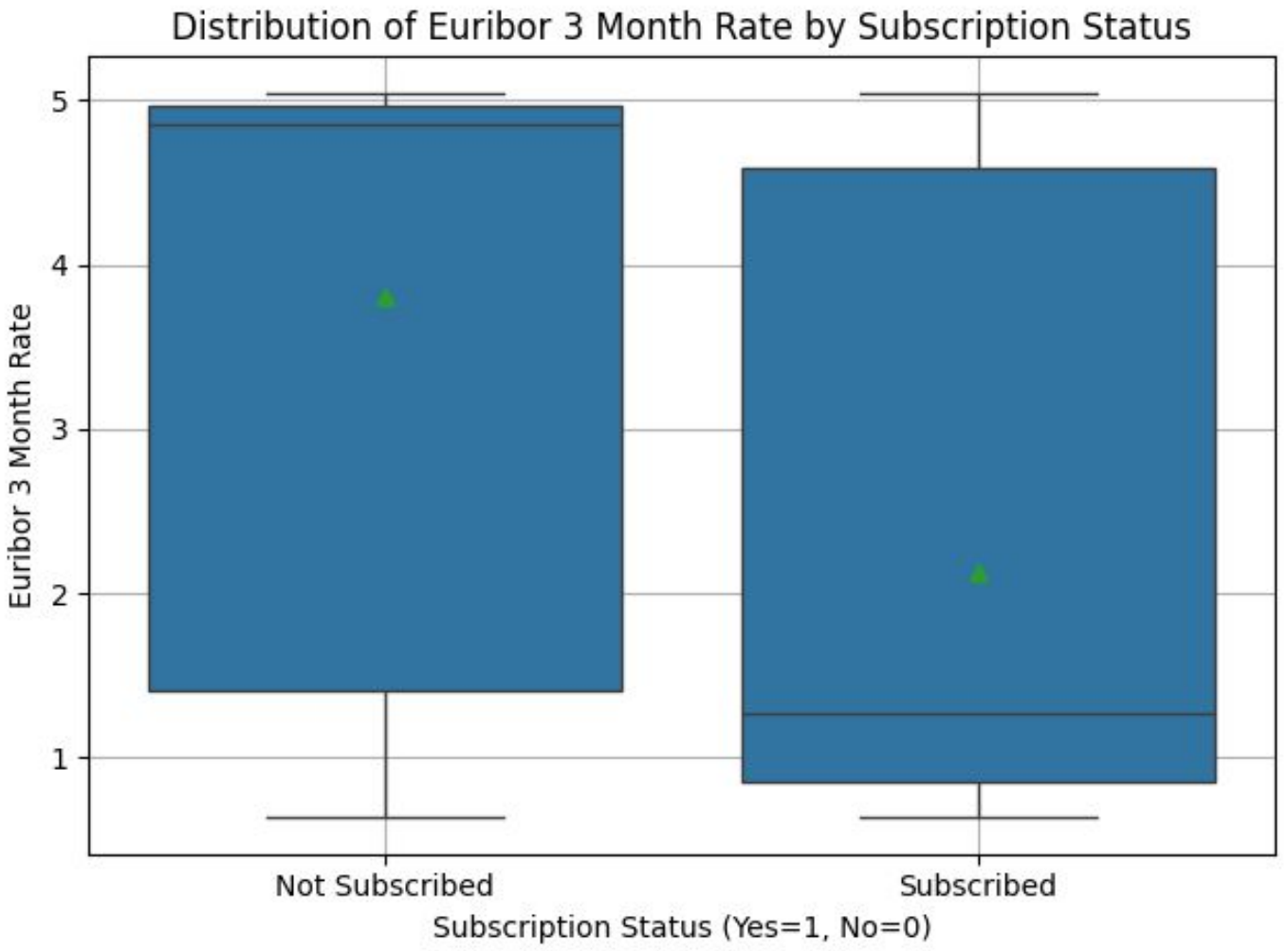
Graphs depict that CCI, CPI and EVR have very weak association. Further investigation needed.



Data Glacier

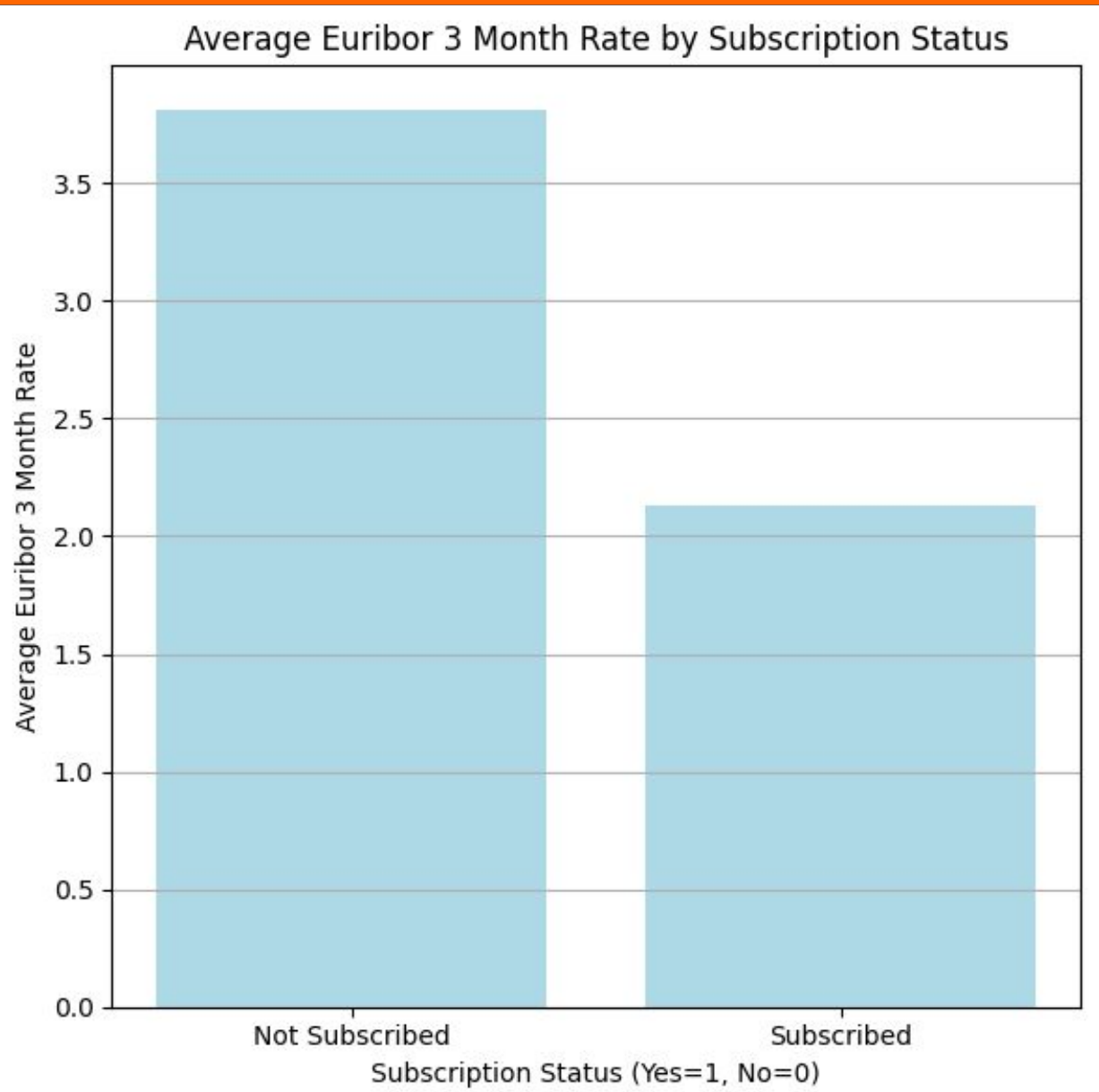
Distribution of Euribor 3 Month Rate by Subscription Status

Customers who subscribed have relatively lower euriborm than those who didn't.



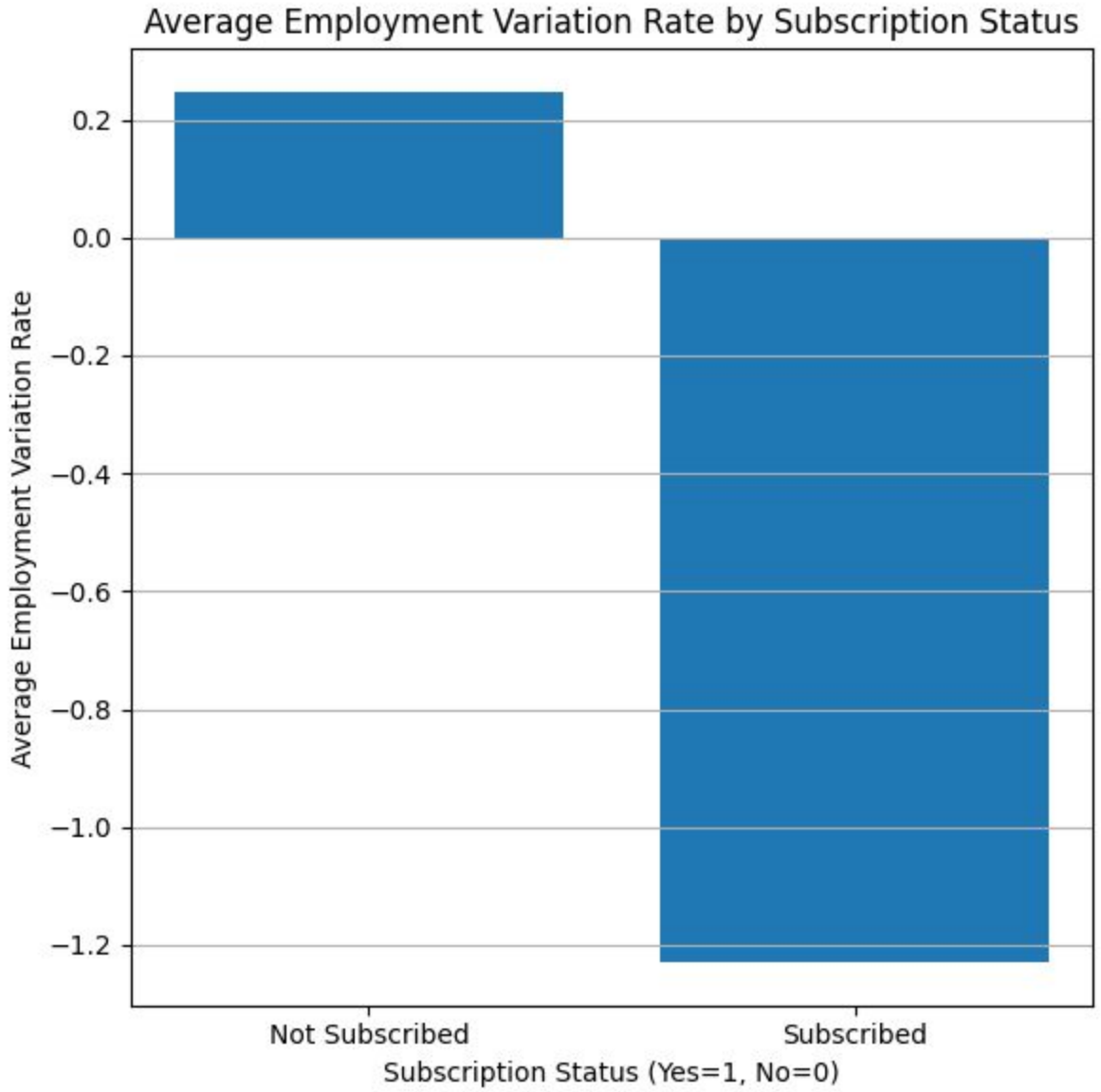
Average Euribor 3 Month Rate by Subscription Status

Customers who fall within the Euribor 3 month rate range of 0.0 to 2.3 might have a higher concentration of subscriptions compared to customers with Euribor 3 month rates outside this range.



Average Employment Variation Rate by Subscription Status

The graphs suggests that customers with lower Emp variation rate are likely to subscribe.

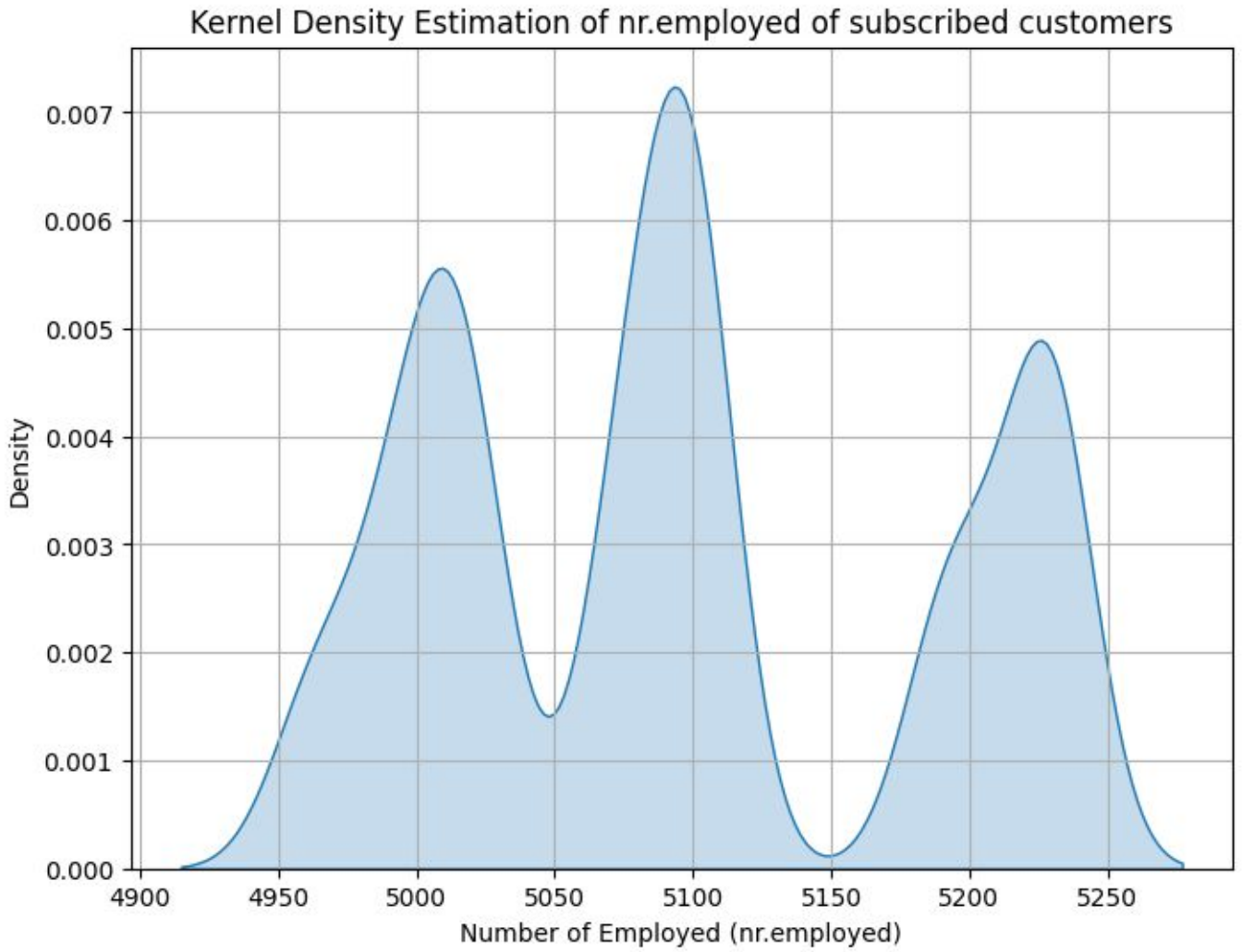


Kernel Density Estimation of nr.employed of subscribed customers

The multimodal distribution of nr.employed indicates the presence of multiple employer categories. This might require using statistical methods suited for non-normal distributions for accurate analysis

Customers' financial conditions does seem to impact their decision.

Hypothesis # 7 Approved.



EDA Summary

- Our exploration of the data has shed light on various factors influencing customer decisions regarding term deposit subscriptions. While specific demographics (retired, mid-aged housemaids), contact methods (cellular), loan status (existing housing loans), seasonal patterns (May weekdays, April Thursdays), and social and economic factors (CPI, CCI, employment variation rates) emerged as key considerations, it's crucial to acknowledge the need for further investigation in all these aspects for a truly comprehensive marketing strategy.

Key Findings :

- An interesting trend emerged from our analysis: a higher proportion of customers who subscribed to term deposits were married, worked in administrative positions, and held university degrees.
- Our analysis suggests a link between customer engagement during the initial calls of the current campaign and their likelihood of subscribing to a term deposit. Customers who demonstrated interest or responded positively during these first interactions seemed more receptive to the product.
- Our exploration of the data revealed interesting patterns in customer demographics and their likelihood to subscribe to term deposits. Customers who were retired or in the older age bracket, followed by those in mid-age and working as housemaids, showed a higher propensity to subscribe compared to other demographics.
- Our analysis revealed an interesting trend regarding the preferred contact method for term deposit subscriptions. Customers reached via cellular phone calls had a higher subscription rate compared to those contacted through traditional telephone calls.
- An interesting insight emerged from our data analysis: customers with existing housing loans were more likely to subscribe to term deposits compared to those without housing loans.

Recommendations :

- Tailoring marketing campaigns to target this specific demographic (married, admin, university degree) as they seem to be more receptive to term deposit products.
- Prioritize outreach to customers who actively engaged during the initial calls (e.g., asking questions, and expressing interest), the bank can focus its resources on those most likely to convert.
- Tailoring messaging and communication strategies to resonate with the needs and interests of retirees, older adults, and middle-aged housemaids can potentially increase the effectiveness of the marketing efforts.
- ABC Bank might consider prioritizing cellular communication methods within its marketing campaigns.
- ABC Bank could consider incorporating a targeted marketing strategy for customers with existing housing loans
- Prioritizing weekday outreach during May and Thursdays in April could potentially lead to higher conversion rates.

Conclusion

- By continuously investigating these areas and integrating the findings into your marketing strategy, ABC Bank can develop a future-proof approach that resonates with customers and achieves optimal results. A data-driven and adaptable approach that prioritizes responsible lending practices and ethical communication will ultimately lead to a successful marketing campaign.

Recommended Models for Dataset:

- Logistic Regression (Base Model)
- Support Vector Machine
- Decision Tree
- Random Forest
- XGBoost
- LightGBM

Thank You