Data Glacier

Your Deep Learning Partner

# Bank Marketing Campaign

Name: Vijayarajan Vijaya Jothi
Location: United Kingdom
Project: Data Science
Date:22/11/2024

# Agenda

- Executive Summary
- Problem Statement
- Approach
- EDA
- EDA Summary
- Recommendations

Data Glacier
Your Deep Learning Partner

Data Glacier
Your Deep Learning Partner

## Executive Summary

### Problem Statement:

- ABC Bank wants to sell its term deposit product to customers and before launching the product they want to develop a model which help them in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution).

### ML Problem:

- With an objective to gather insights on the factors that are impacting the persistency, build a classification for the given dataset.

The highest model accuracy and precision were attained using the Random Forest Classifier model.

Data Glacier

Your Deep Learning Partner

Problem Statement

## Business Understanding:

Bank wants to use ML model to shortlist customer whose chances of buying the product is more so that their marketing channel (tele marketing, SMS/email marketing etc)  can focus only to those customers whose chances of buying the product is more. This will save resource and their time  (which is directly involved in the cost ( resource billing)).Develop model with Duration and without duration feature and report the performance of the model. Duration feature is not recommended as this will be difficult to explain the result to business and also it will be difficult for business to campaign based on duration.

Data Glacier

Your Deep Learning Partner

## Approach

1. Business Understanding
2. Importing Required libraries and dataset
3. Exploratory Data Analysis
4. Data processing and transformation
5. Model Building
6. Model evaluation
7. Model Deployment

# Exploratory Data Analysis
# Bank Marketing Campaign

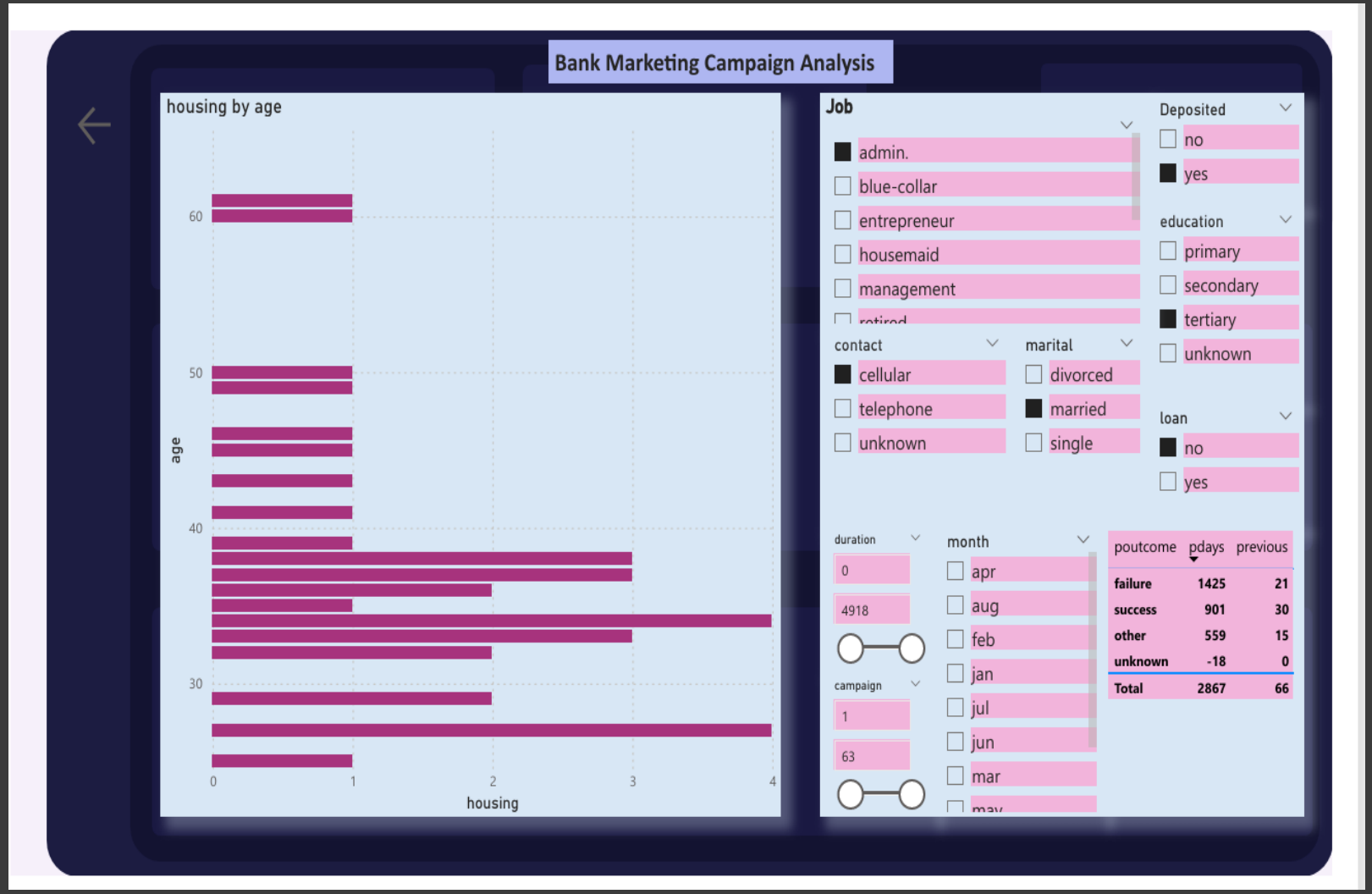
- File Used: bank-additional-full.csv
- Correlation between all variables and the predictor.
- Missing data or nulls exist and have been handled.
- Data wrangling transformation included normalizing data and standardize them.
- This has increased the correlation between the features and the predictor variable.
- Dummy variables have been created (Categorical variables to 0 and 1)

# Exploratory Data Analysis
# Bank Marketing Campaign

- File Used: bank-additional-full.csv
- Visual elements in the Dashboard gives insights about Customer base.
- Pie Chart shows Majority Customers having housing have obtained Secondary Education .
- Combo chart visualizations shows Customers with Secondary Education having higher balance and Deposits in Bank.
- From Charts, “May”, Month remain the highest  Loan applied.
- Also, Customers between Age 20 and 80 having housing ownership.

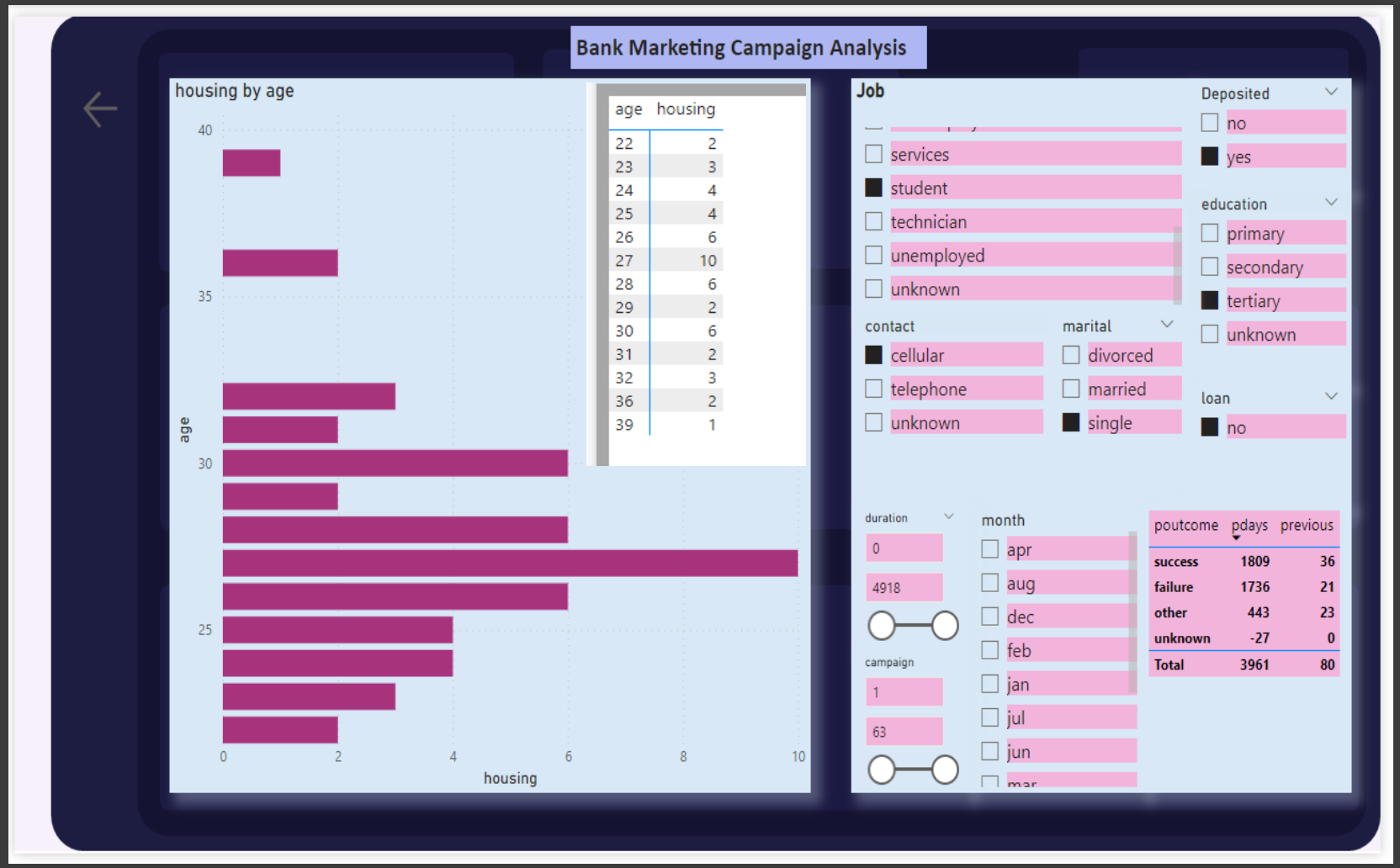
# Exploratory Data Analysis
# Bank Marketing Campaign

- File Used: bank-additional-full.csv
- Visual elements in the Dashboard gives insights about Customer base.
- From charts, Customers in classes of admin, cellular, and married , having deposits and possess no loan and obtained tertiary education, Got Housing ownership between 20 and 60 .
- From Charts, Customers between age 30 to 40 falling in the above filtered conditions, have maximum housing ownership.



# Exploratory Data Analysis
# Bank Marketing Campaign

- File Used: bank-additional-full.csv
- Visual elements in the Dashboard gives insights about Customer base.
- From charts, Customers in classes of student, cellular, and single , having deposits and possess no loan and obtained tertiary education, Got Housing ownership between 22 and 39 .
- From Charts, Customers between age 23 to 32 falling in the above filtered conditions, have maximum housing ownership.

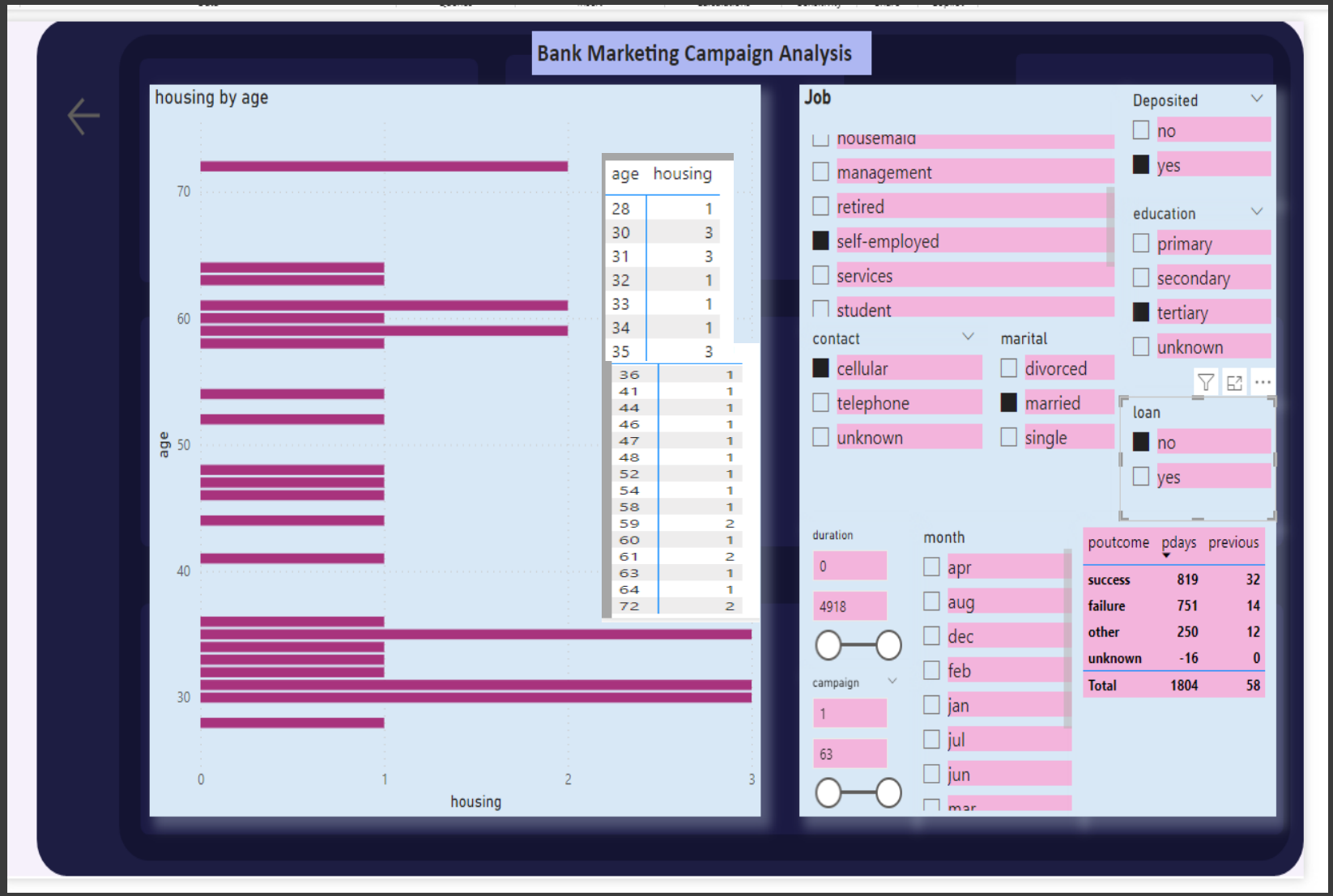


# Exploratory Data Analysis
# Bank Marketing Campaign

- File Used: bank-additional-full.csv
- Visual elements in the Dashboard gives insights about Customer base.
- From charts, Customers in classes of self-employed, cellular, and married , having deposits and possess no loan and obtained tertiary education,

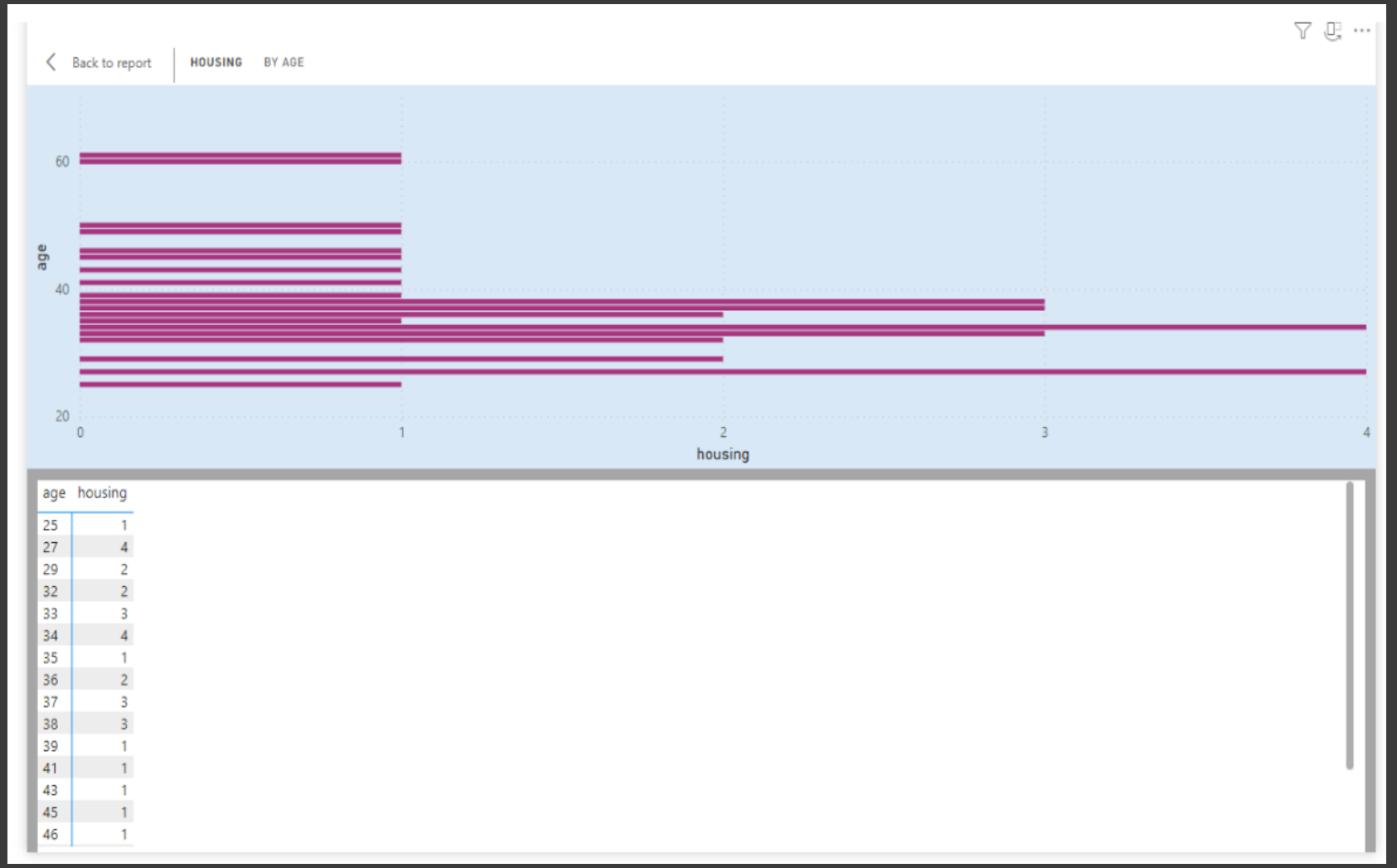
Got Housing ownership between 28 and 72 .

- From Charts, Customers between age 28 to 35 falling in the above filtered conditions, have maximum housing ownership.



# Exploratory Data Analysis
# Bank Marketing Campaign

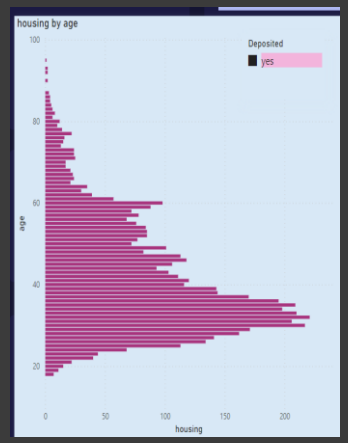
- File Used: bank-additional-full.csv
- Visual elements in the Dashboard gives insights about Customer base.
- From charts, Customers in classes of admin, cellular, and married , having deposits and possess no loan and obtained tertiary education, Got Housing ownership between 20 and 60 .
- From Charts, Customers between age 30 to 40 falling in the above filtered conditions, have maximum housing ownership.

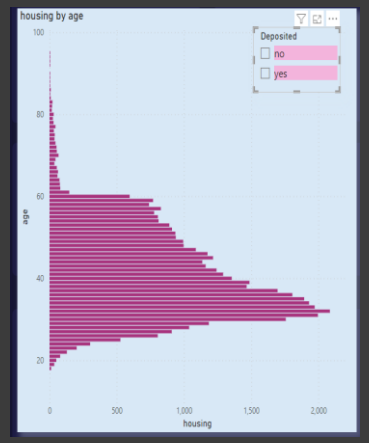


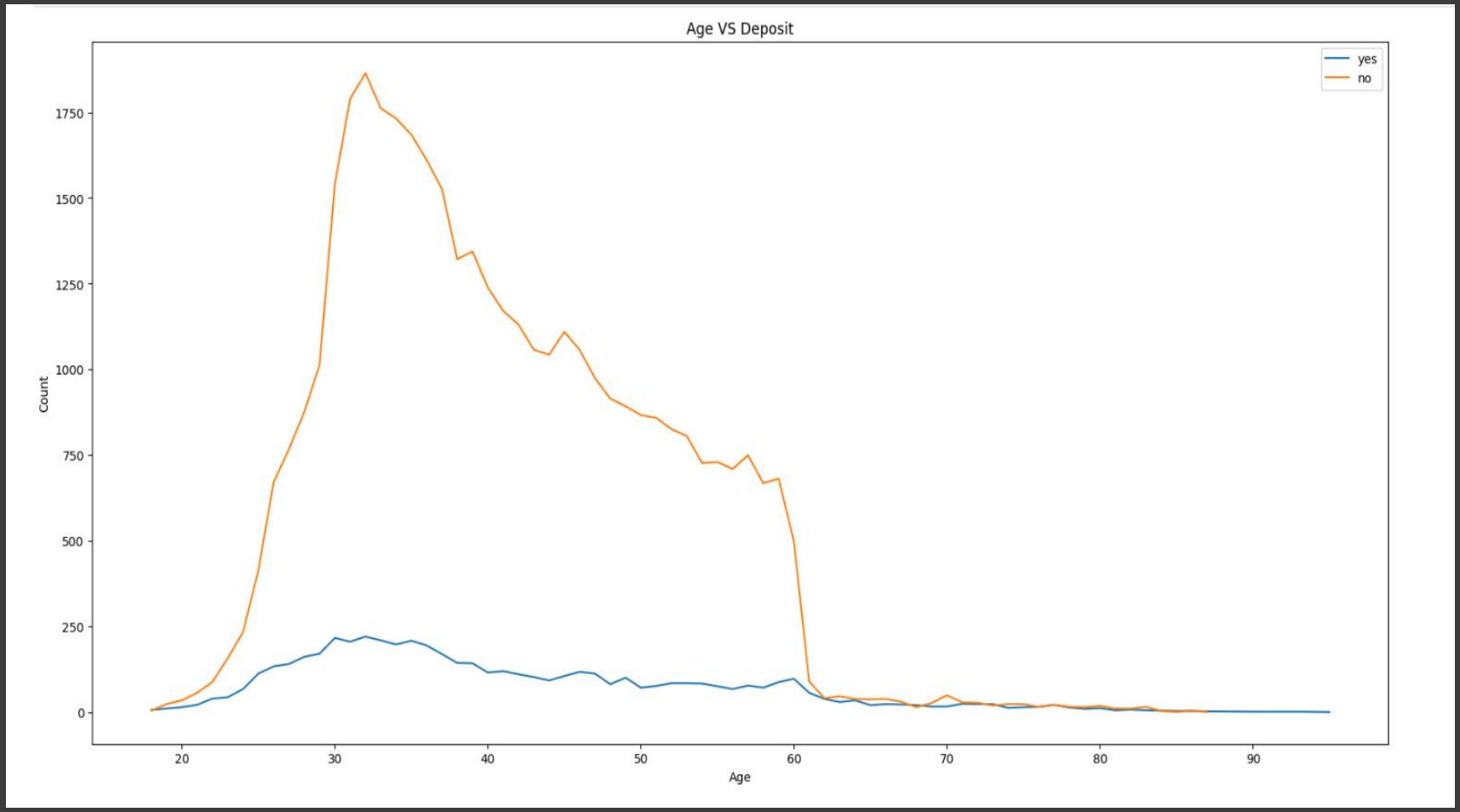

| age | housing |
|---|---|
| 25 | 1 |
| 27 | 4 |
| 29 | 2 |
| 32 | 2 |
| 33 | 3 |
| 34 | 4 |
| 35 | 1 |
| 36 | 2 |
| 37 | 3 |
| 38 | 3 |
| 39 | 1 |
| 41 | 1 |
| 43 | 1 |
| 45 | 1 |
| 46 | 1 |

# Exploratory Data Analysis
# Bank Marketing Campaign

- File Used: bank-additional-full.csv
- The line graph shows customers between Age 20 and 60 have Deposits in Bank.



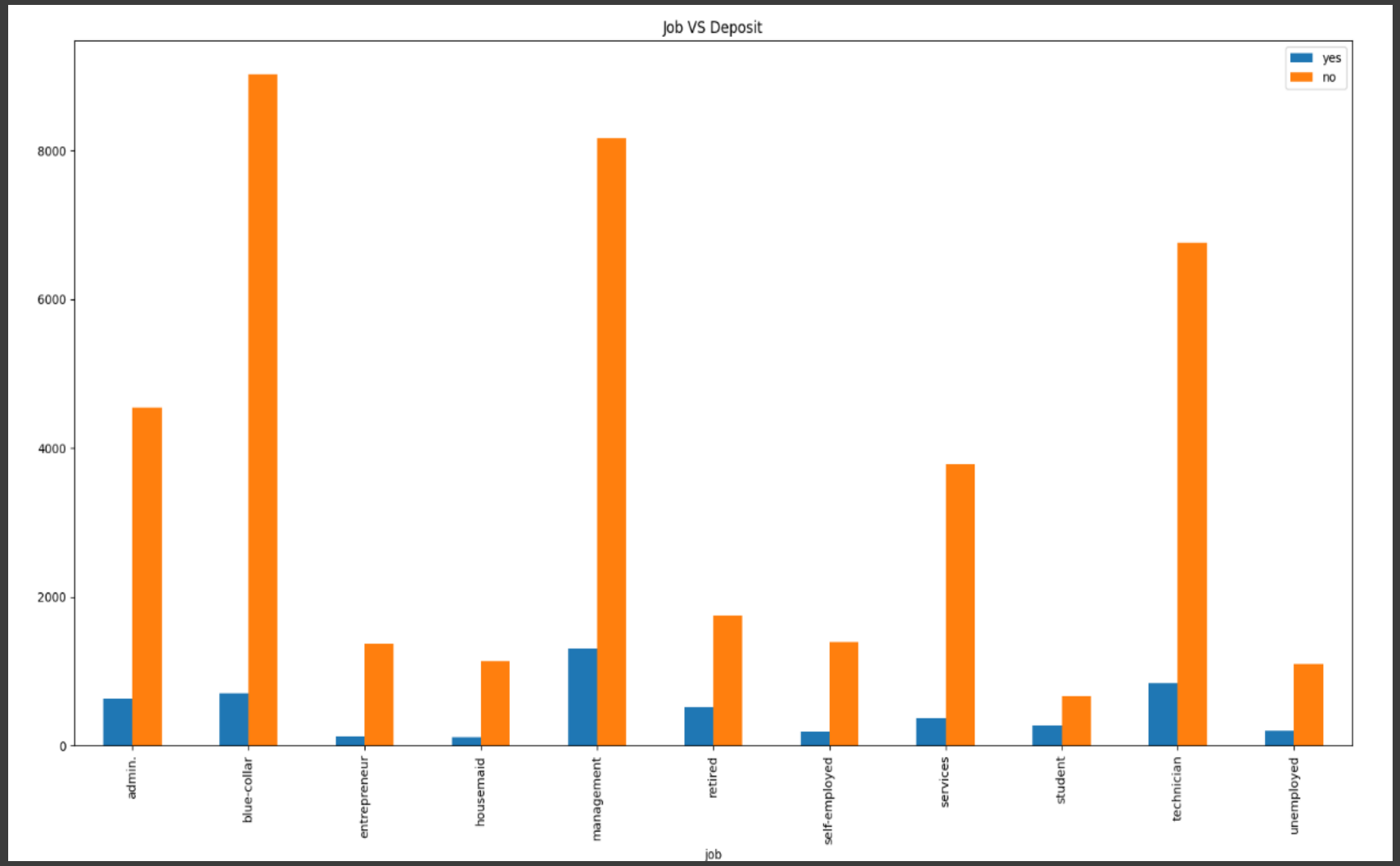



# Exploratory Data Analysis
# Bank Marketing Campaign

- File Used: bank-additional-full.csv
- The Stacked Column Chart shows Customers having Deposit based on job.

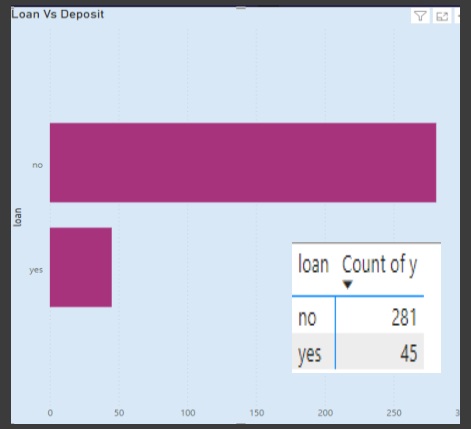
From charts, it is explained that customer having management jobs have comparatively got higher Deposits in Bank.

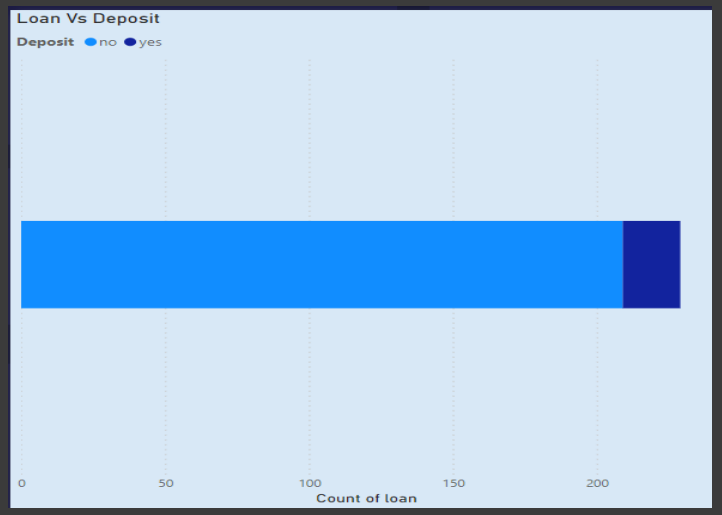


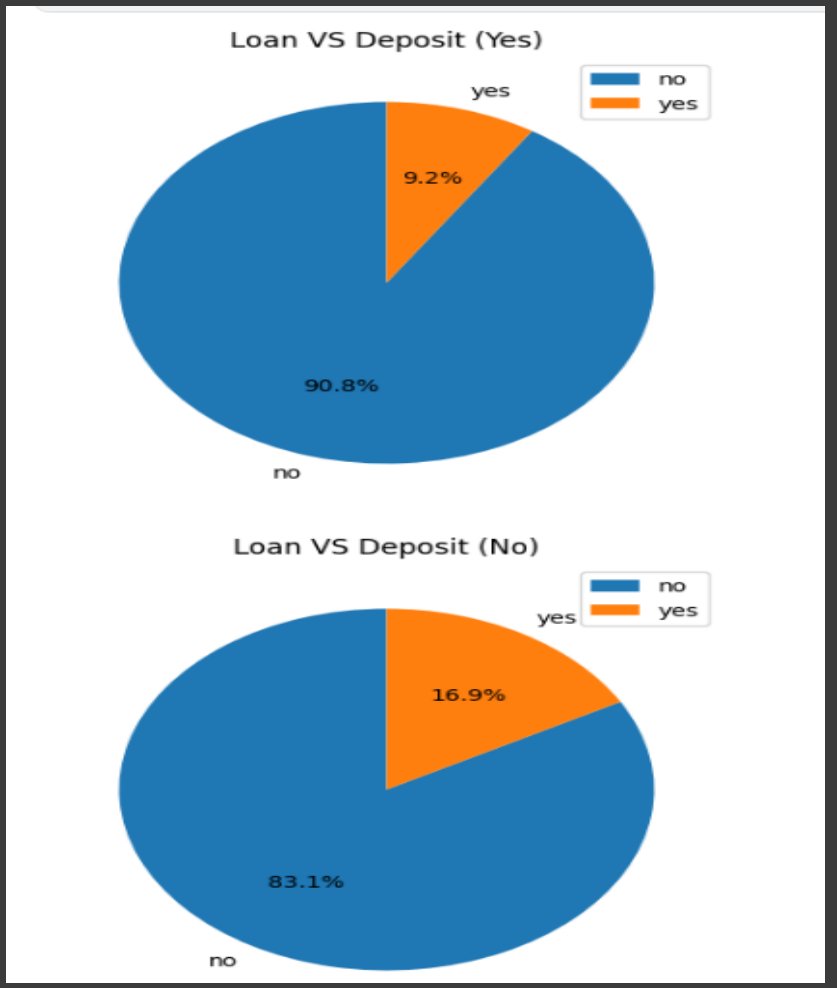
# Exploratory Data Analysis
# Bank Marketing Campaign

- File Used: bank-additional-full.csv
- The First Pie chart shows 9.2% of customers with Deposit have loan in bank

- File Used: bank-additional-full.csv
- The Second Pie chart shows 16.9% of customers with loan have no Deposit in bank



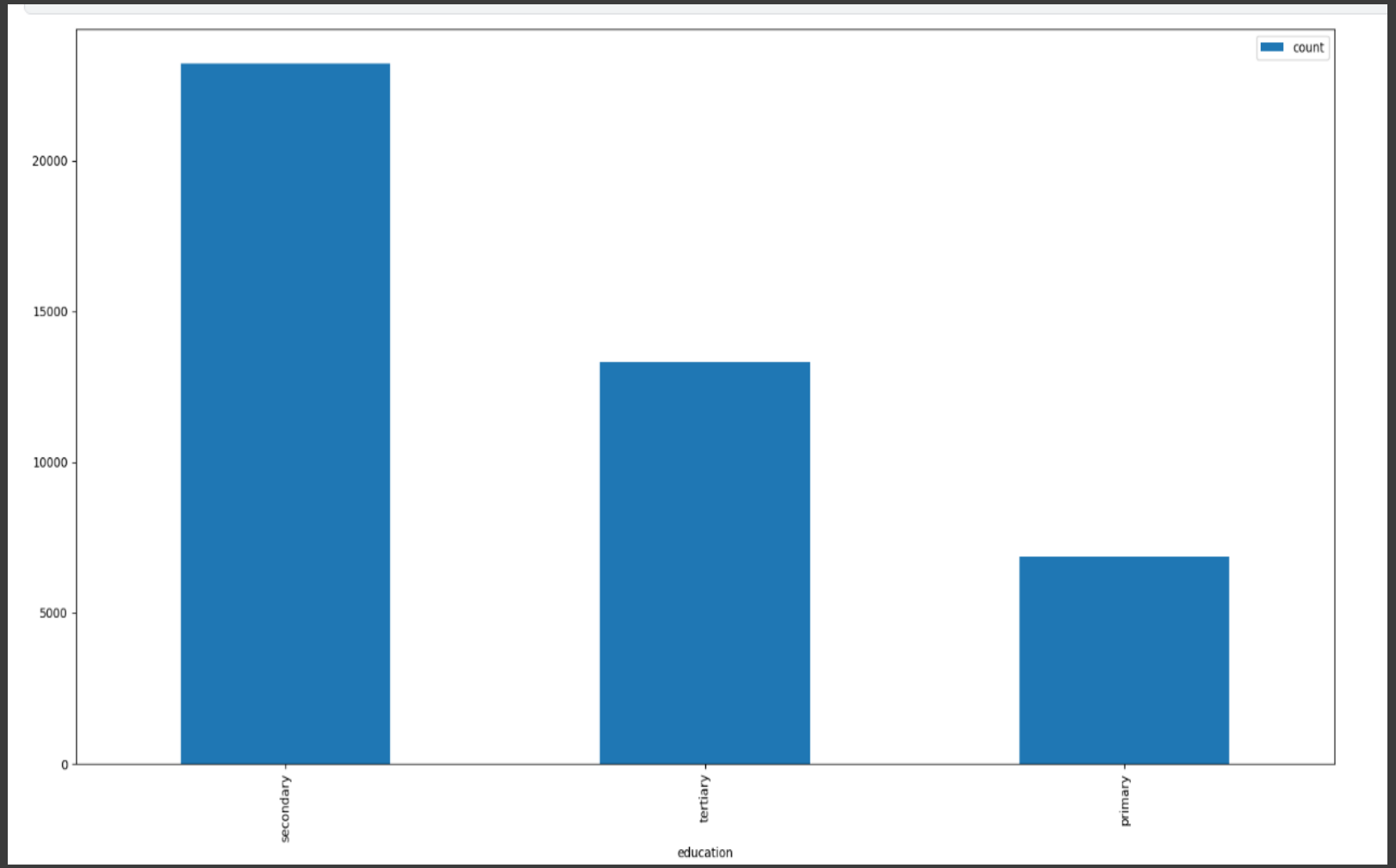



# Exploratory Data Analysis
# Bank Marketing Campaign

- File Used: bank-additional-full.csv

- The Column Chart shows Customers having Secondary Education have higher deposits in Bank



# Exploratory Data Analysis
# Bank Marketing Campaign

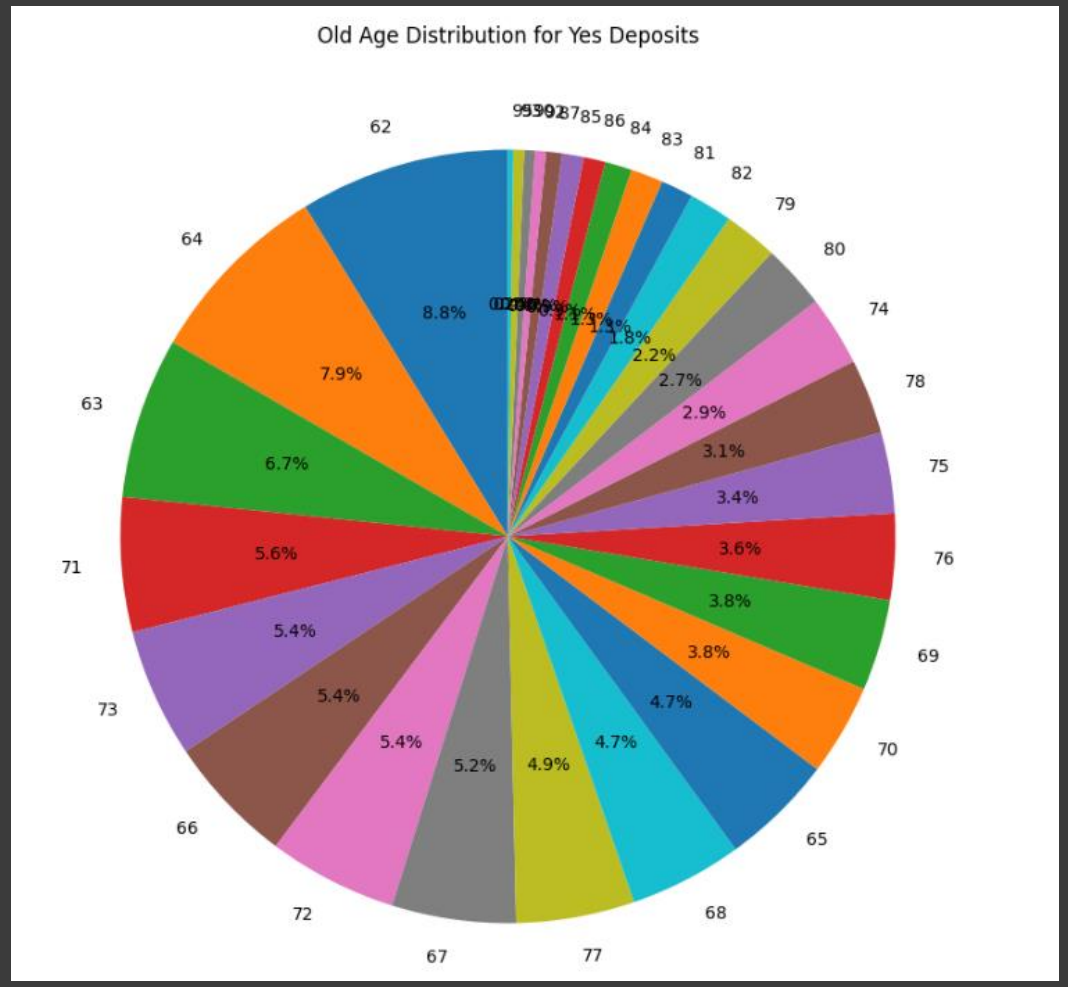
```
import matplotlib.pyplot as plt


yes_old_age = old_age[old_age['y'] == 'yes']


yes_old_age['age'].value_counts().plot.pie(autopct='%1.1f%%', startangle=90, figsize=(10, 20))
plt.title('Old Age Distribution for Yes Deposits')
plt.ylabel('')
plt.show()
```

- File Used: bank-additional-full.csv

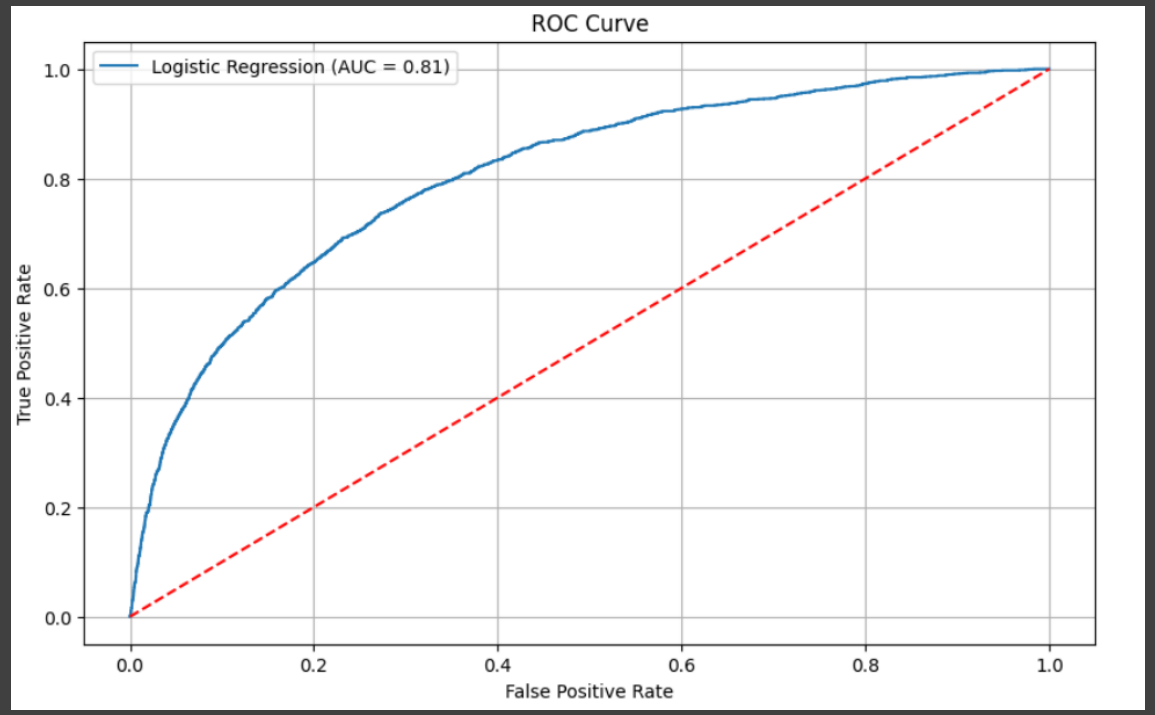
•The Pie Chart shows Distribution of Old Age Customers having deposits in Bank.
About 8.8% of customers in Old Age 62 have max Deposits in Bank .



# Model Building

- The case is classification so we will be using three models:
- Logistic Regression,
- Ridge Regression,
- Random Forest classifier.

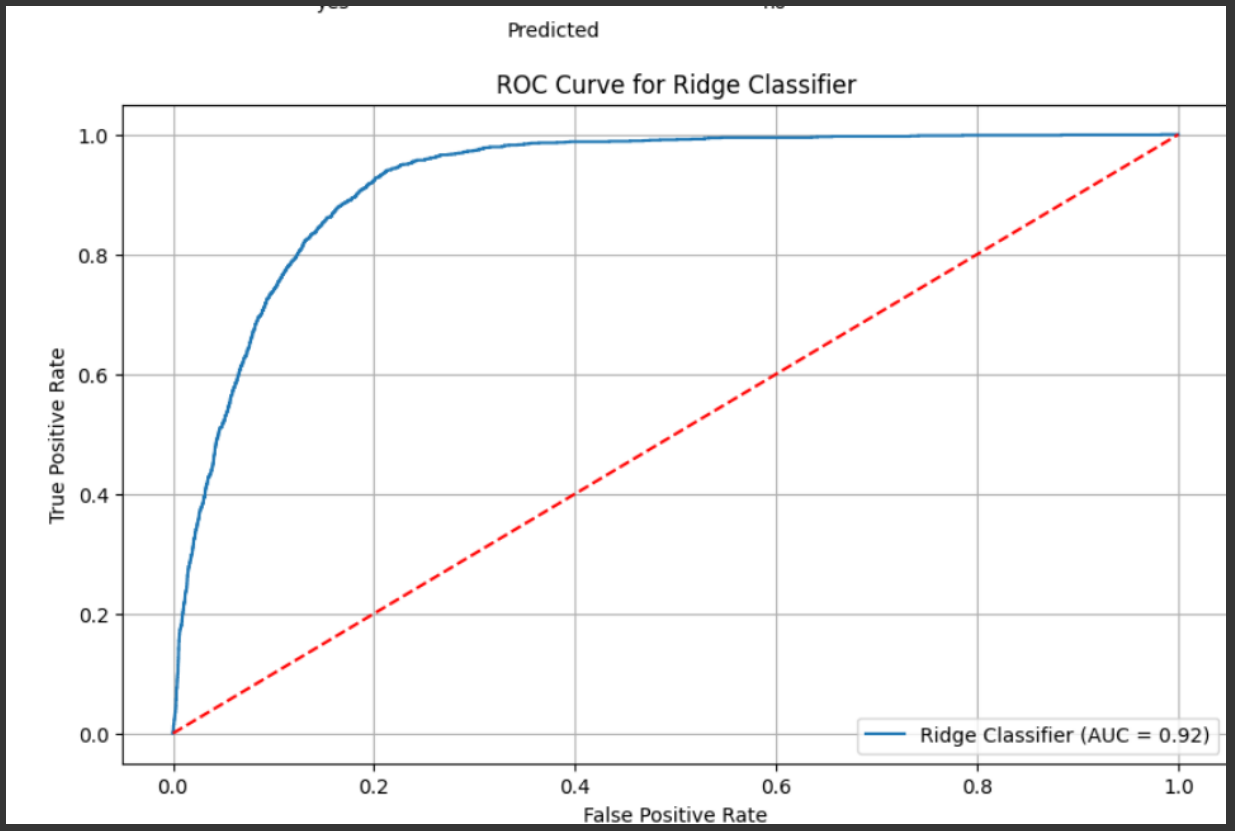
# Logistic Regression Model Results




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| No Deposited | 0.91 | 0.97 | 0.94 | 11977 |
| Deposited | 0.55 | 0.25 | 0.34 | 1587 |
| accuracy | | | 0.89 | 13564 |
| macro avg | 0.73 | 0.61 | 0.64 | 13564 |
| weighted avg | 0.87 | 0.89 | 0.87 | 13564 |

Accuracy: 0.8883810085520495
AUC: 0.8054506276706893

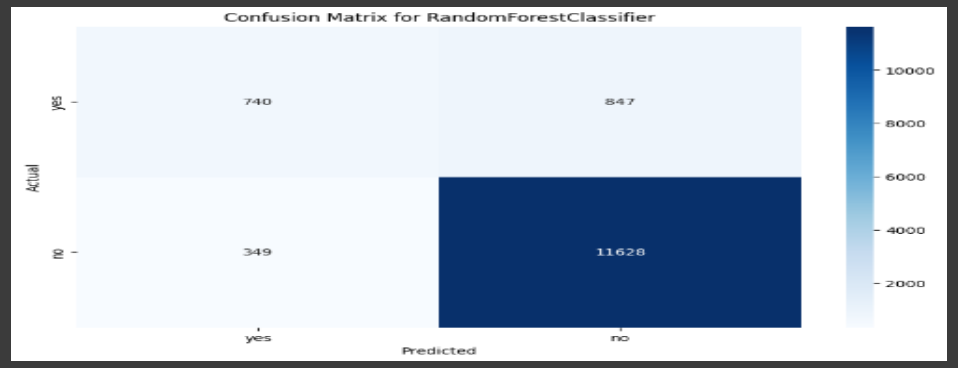
# Ridge Regression Model Results

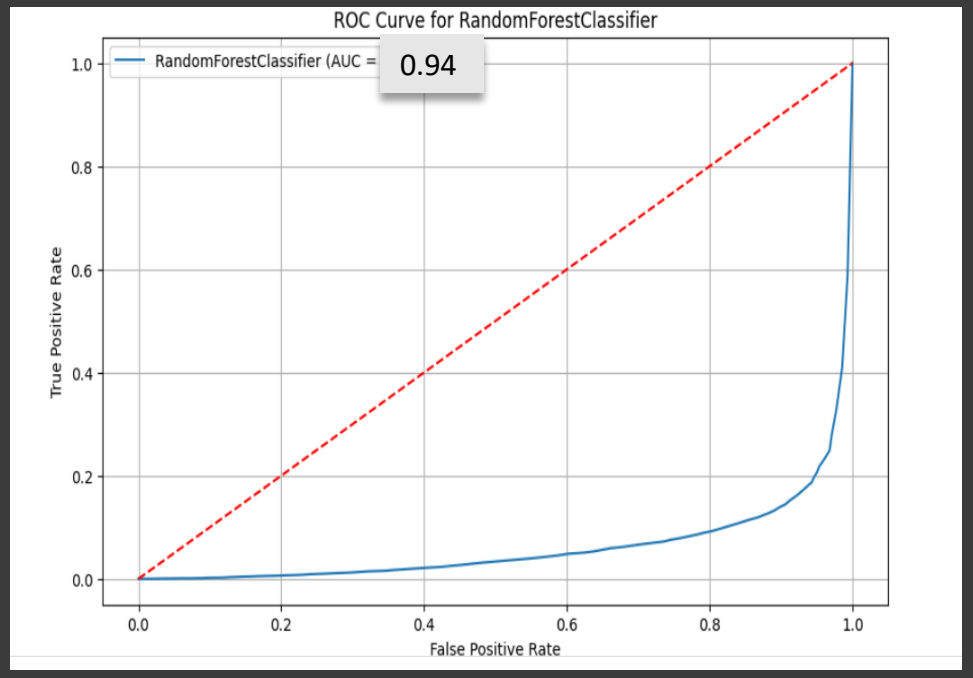



| | precision | recall | f1-score | s |
|---|---|---|---|---|
| no | 0.91 | 0.98 | 0.95 | |
| yes | 0.68 | 0.30 | 0.42 | |
| accuracy | | | 0.90 | |
| macro avg | 0.80 | 0.64 | 0.68 | |
| weighted avg | 0.89 | 0.90 | 0.88 | |

Accuracy: 0.9012828074314362
Precision: 0.6766381766381766
Recall: 0.29930686830497794
F1 Score: 0.4150283966797728
AUC: 0.9243818979025068

# Random Forest Classifier Model Results




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| No Deposited | 0.93 | 0.97 | 0.95 | 11977 |
| Deposited | 0.68 | 0.47 | 0.55 | 1587 |
| accuracy | | | 0.91 | 13564 |
| macro avg | 0.81 | 0.72 | 0.75 | 13564 |
| weighted avg | 0.90 | 0.91 | 0.90 | 13564 |

Accuracy: 0.9118254202300207

AUC: 0.9380842266518072

# Random Forest Model

ModelTrade-offs:

Advantages:

- Insensitive to Outliers.
- Insensitive to Null values.
- Less Prone to overfitting.

Disadvantages:

- Losing Interpretability.
- Difficult to diagnose and improve.

Results obtained:

- Accuracy: 87– 91 %

# Conclusion

Approximately all the classifiers have same result, but Random Forest was the best one.
The model has around 91% Accuracy.
Random Forest has 93% Precision, 95% Recall, & 95% F1 Score.
We can also see the results for each classifier as well

# Thank You