**Results of Analysis**

**Results after EDA using SQL:**Here, we considered various parameters to analyse how the different parameters influenced the final outcome , i.e, did the candidate take up the term deposit? Or Not?

Lets look at all the input variables present and what they represent.

**Input variables:**

# bank client data:

1 - age (numeric)

2 - job : type of job (categorical: "admin.","blue-collar","entrepreneur","housemaid","management","retired","self-employed","services","student","technician","unemployed","unknown")

3 - marital : marital status (categorical: "divorced","married","single","unknown"; note: "divorced" means divorced or widowed)

4 - education (categorical: "basic.4y","basic.6y","basic.9y","high.school","illiterate","professional.course","university.degree","unknown")

5 - default: has credit in default? (categorical: "no","yes","unknown")

6 - housing: has housing loan? (categorical: "no","yes","unknown")

7 - loan: has personal loan? (categorical: "no","yes","unknown")

# related with the last contact of the current campaign:

8 - contact: contact communication type (categorical: "cellular","telephone")

9 - month: last contact month of year (categorical: "jan", "feb", "mar", ..., "nov", "dec")

10 - day\_of\_week: last contact day of the week (categorical: "mon","tue","wed","thu","fri")

11 - duration: last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y="no"). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

# other attributes:

12 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)

13 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)

14 - previous: number of contacts performed before this campaign and for this client (numeric)

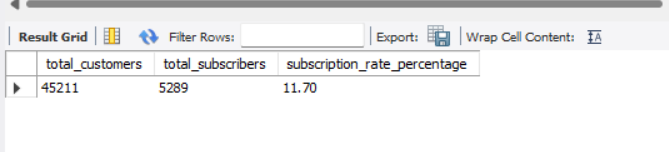
15 - poutcome: outcome of the previous marketing campaign (categorical: "failure","nonexistent","success")

We conducted analysis on all the major ways the input variable would influence the outcome. We checked areas like :-

* Total customers vs total subscribed
* subscription rate per education
* subscription rate per job
* Subscription rate per age group
* Subscription rate per marital status
* subscription rate per personal loan
* subscription per housing loan
* avg\_call\_duration

Lets take a look at the results of the EDA with respect to each scenario:

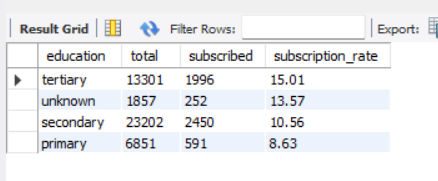
**Total customers vs total subscribed:**



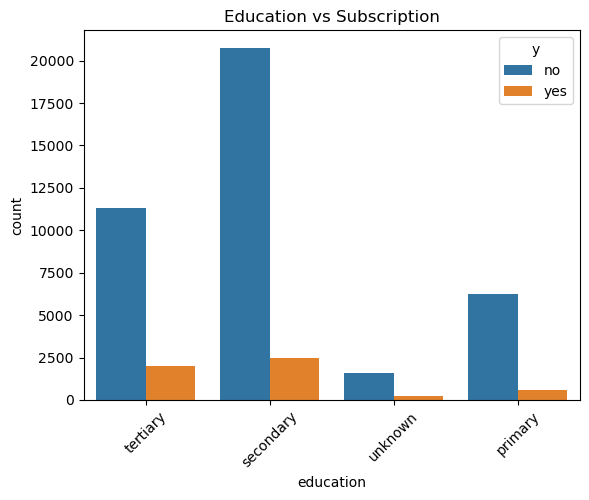
Here, we can see that only 11.70 percent of people actually subscribed from the entire customer pool.

**Subscription rate per education:**

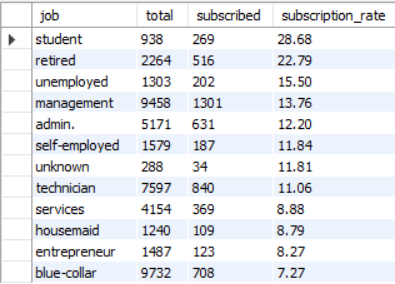
Lets analyse how the education of a candidate influences his term deposit taking capability



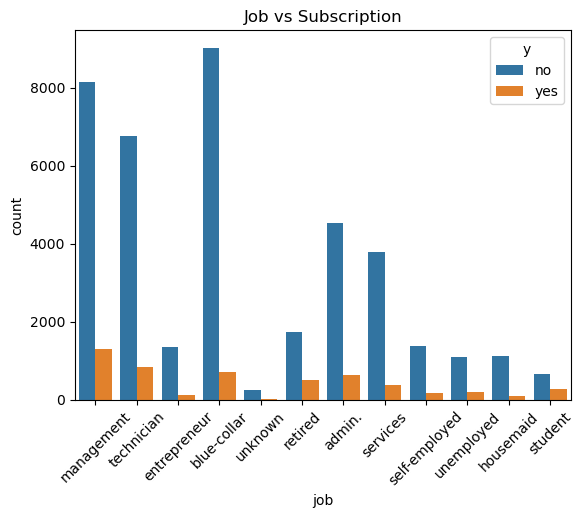
We understand that the higher the education level of the candidate is, the more likely they are to take a term deposit. Its also influenced by many other factors



**Subscription rate per job: -**



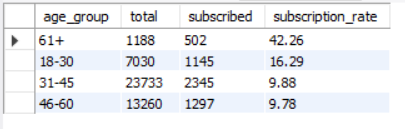
Here, we see an interesting pattern, The highest job group of people who signed up for a term deposit were students and retirees. And Blue collar workers were the least likely to sign up for a term deposit.

This points to multiple interesting insights, The fact that students subscribed most to the term deposits means that the bank had options for small amount term deposits which could be feasible for the students. Students might be financially cautious and saving up their scholarships / internship money .Also it points to the fact that people with lesser financial liability (retirees and students) are more susceptible to subscription to a term Loan. 

We can observe the graph for more insights.

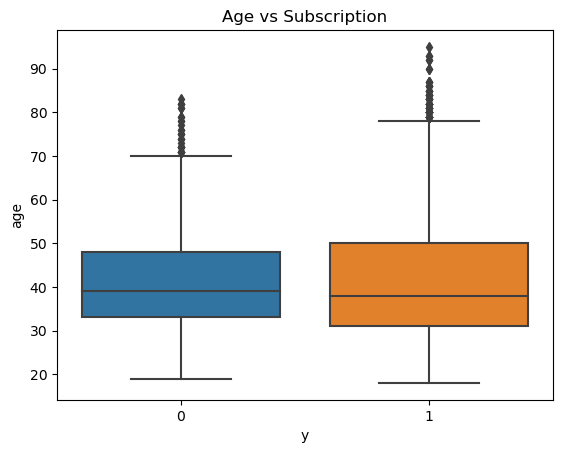
**Subscription rate per age group :-**

Lets analyse the subscription rate per age group: -

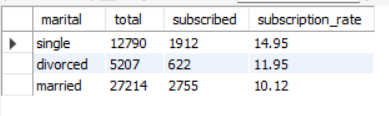


This is again surprising, because we analysed Students to be the highest demographic to subscribe , yet the highest age group to have subscribed is retirees ( we are assuming they are retirees since they are all 61+)

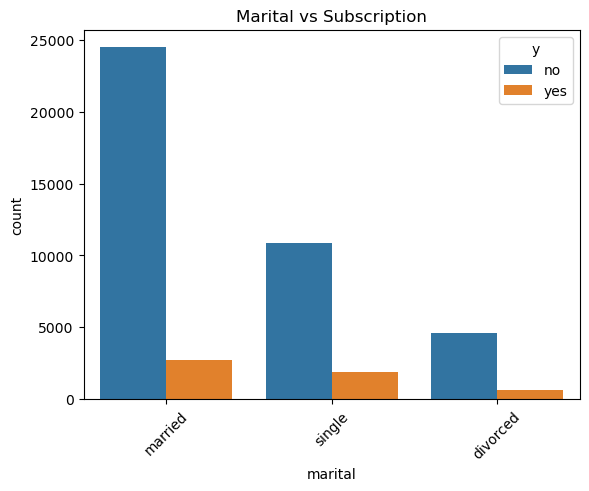
We can infer from this data that the campaign marketed less number of students however got the highest subscription rate per job title from them. However since the gross subscription rate seems to be highest for retirees , we can infer more retirees were targeted in the campaign. Possibly because retirees have a higher chance of depositing a sizeable fund to the term deposit.



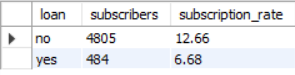
**Subscription rate per marital status :-**

****

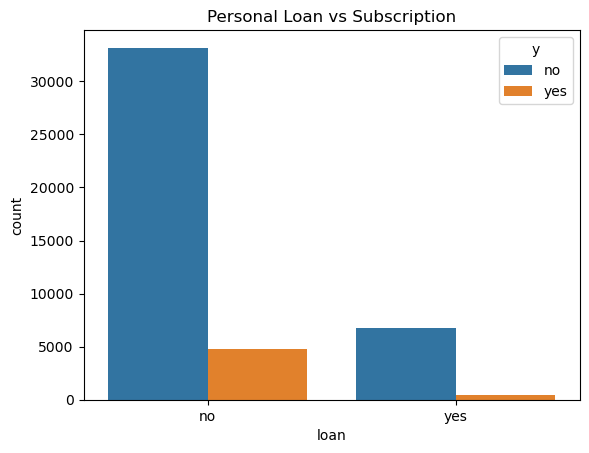
Unsurprisingly, single people were more susceptible to a term deposit subscription since they have a lesser financial burden



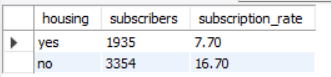
**Subscription rate per personal loan :-**



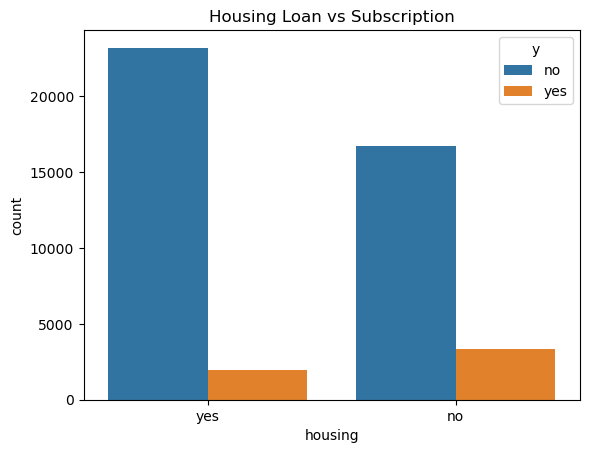
We can notice that a personal loan directly impacts term deposit subscription rates. People without the burden of a personal loan are more susceptible to subscribe to a term deposit



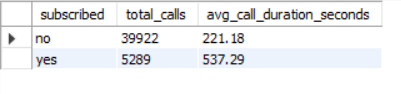
**Subscription per housing loan :-**



Same as personal loan, people without the burden of a housing loan are more likely to subscribe , however we see more people with house loan subscribing than people with personal loan. This could be because Housing loans are long term, lower interest rate and stable as opposed to personal loans which are much more volatile and have a higher interest rate which makes it difficult for candidates to save money .

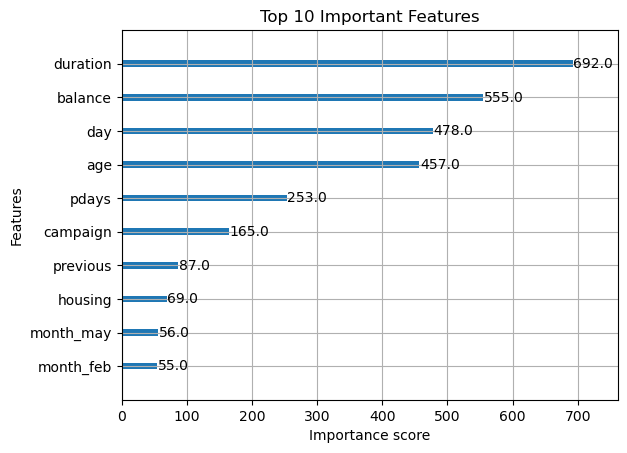


**Subscription per call Duration:-**

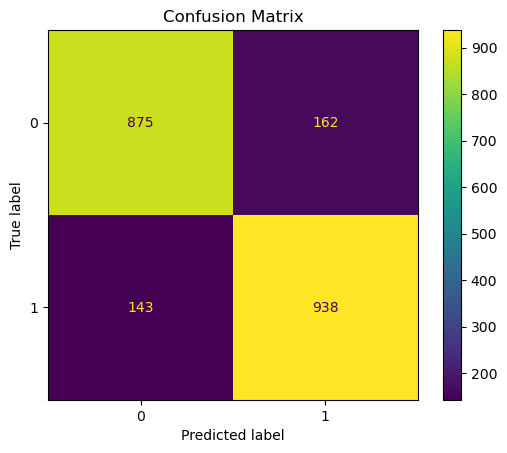


Call duration played a huge role in subscription turnout because we can see higher call duration corelated with positive turnout.

These are the Top features that determined positive turnout according to our Model.



Lets see the performance of our model:



**Understanding the Matrix:**

| **Predicted: No (0)** | **Predicted: Yes (1)** |  |
| --- | --- | --- |
| **Actual: No (0)** | **875** | **162** |
| **Actual: Yes (1)** | **143** | **938** |

**Metrics You Can Derive:**

**True Positives (TP): 938**

* Model correctly predicted **"Yes"** when the actual label was **"Yes"**.

**False Positives (FP): 162**

* Model predicted **"Yes"**, but the actual label was **"No"**.

**False Negatives (FN): 143**

* Model predicted **"No"**, but the actual label was **"Yes"**.

**True Negatives (TN): 875**

* Model correctly predicted **"No"** when the actual label was **"No"**.

**Performance Metrics:**

1. **Accuracy** = (TP + TN) / Total  
   = (938 + 875) / (938 + 875 + 143 + 162) = 1813 / 2118 ≈ 85.6%
2. **Precision (for class 1)** = TP / (TP + FP)  
   = 938 / (938 + 162) ≈ **85.3%**  
   → When the model predicts "Yes", it's correct 85.3% of the time.
3. **Recall (for class 1)** = TP / (TP + FN)  
   = 938 / (938 + 143) ≈ **86.7%**  
   → Out of all actual "Yes", it correctly finds 86.7%.
4. **F1-Score (for class 1)** = 2 × (Precision × Recall) / (Precision + Recall)  
   ≈ 2 × (0.853 × 0.867) / (0.853 + 0.867) ≈ **86%**

Our Model is peforming reasonably well with F1 score for around 86%.