Slot-1 report

Oasis crew team 11

Data Science Problem Statement:

Predict if the client will subscribe to a term deposit based on the analysis of the marketing campaigns the bank performed.

Preprocessing Techniques applied:

a) Missing Value Treatmesnt:

→No null values found in numerical columns and object columns

→"Unknown" values are identified as missing value for categorical columns

→ Mode is used to fill "unknown" of each categorical column

 b) Label Encoding:

→Label encoding on all the categorical features and the target is performed.

c) Imbalanced Data Handling:

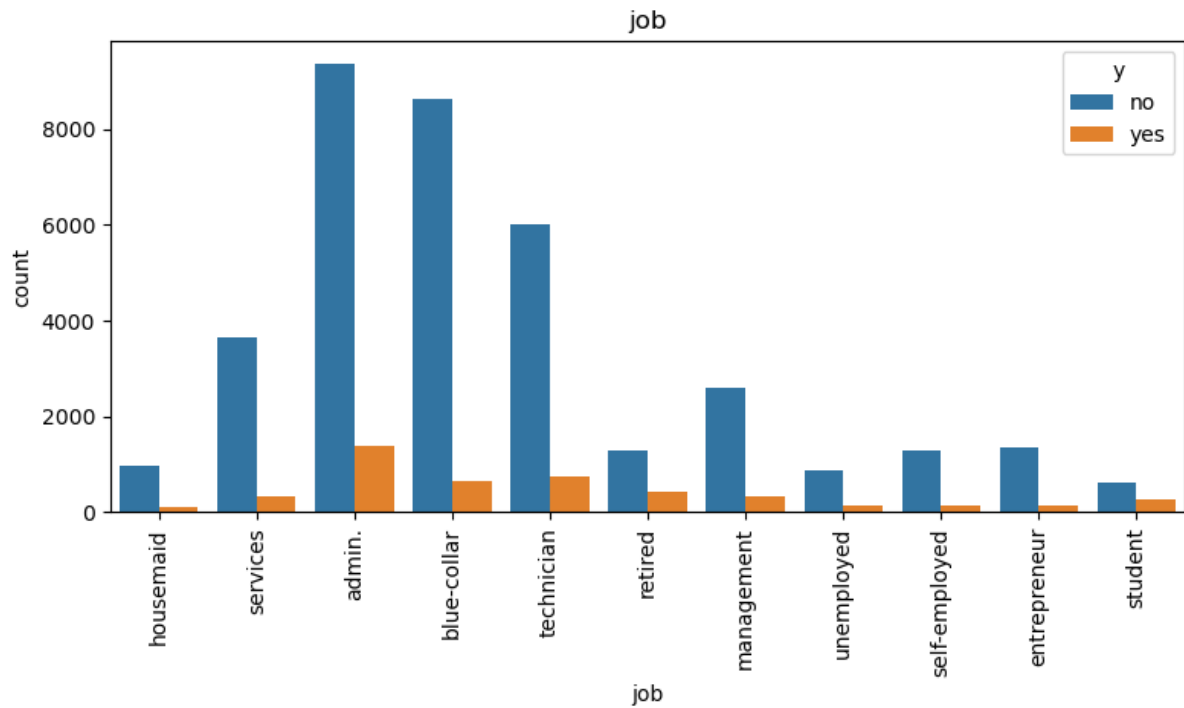→ The class distribution in the target indicates the problem imbalanced.

```
0      88.734583
1      11.265417
Name: y, dtype: float64
```

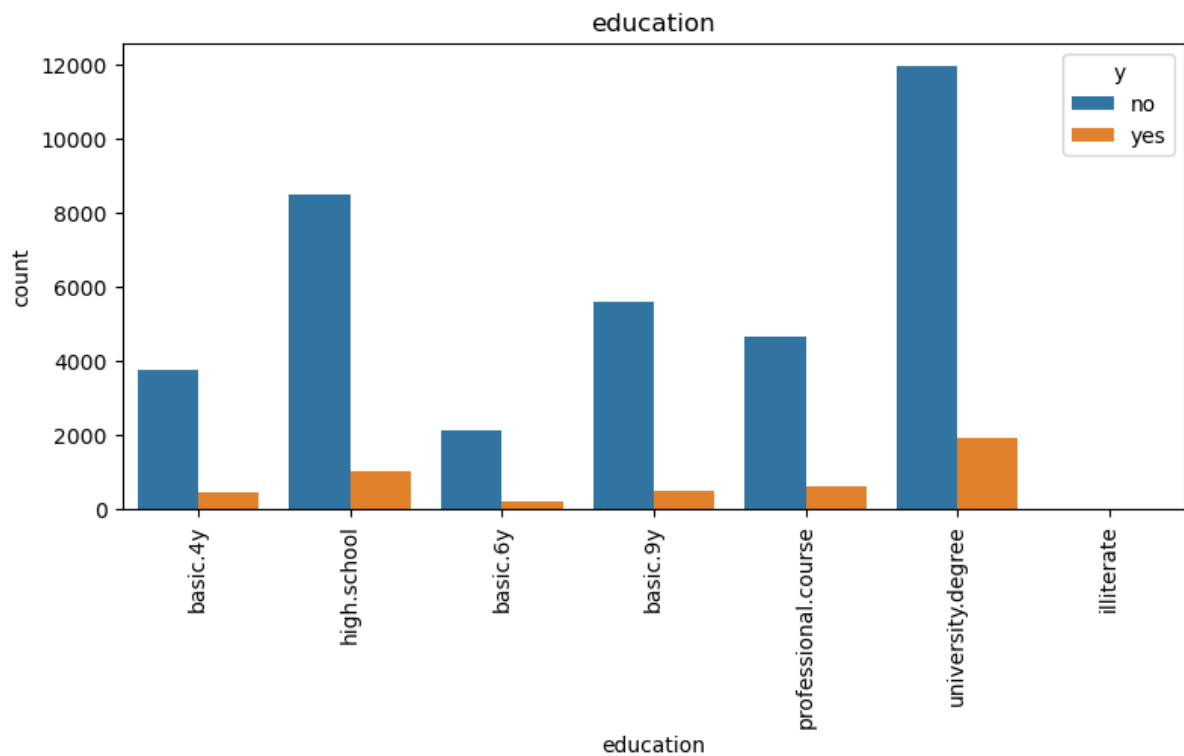## Exploratory Data Analysis:

Count plots:

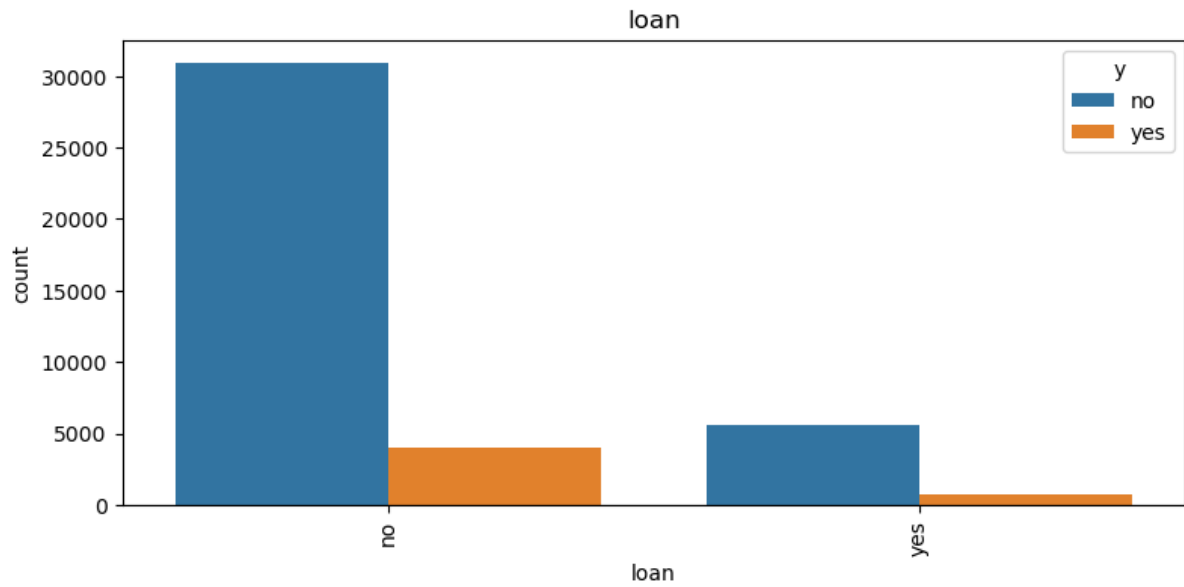Shows the trends of some categorical columns.

- -Job - Admin, Blue-collared and Technician are most likely to subscribe to a term deposit.
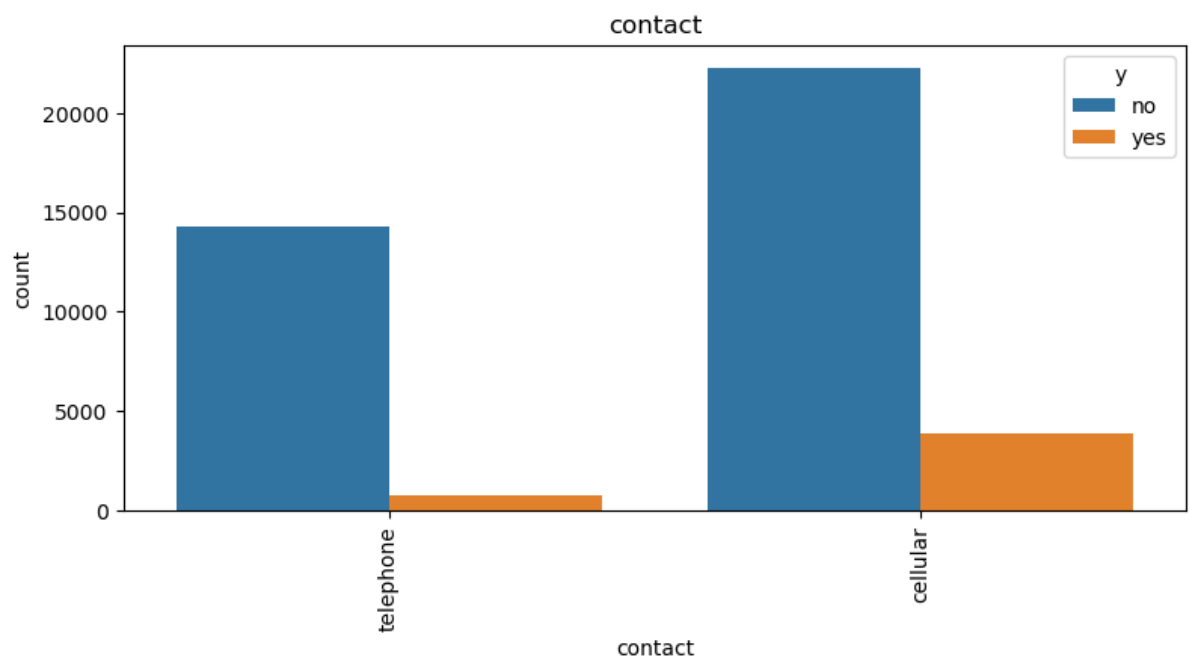
- -Education - People from education sector(university and high school)are most likely to subscribe to a term deposit.
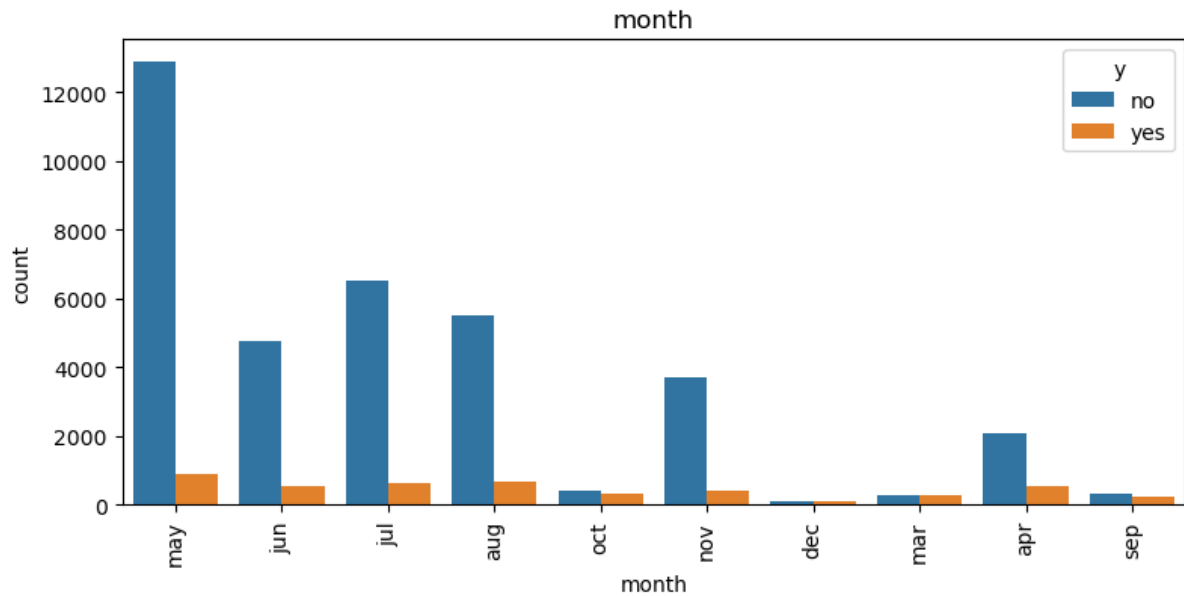

- -Loan - People with no personal loan are most likely to subscribe to a term deposit.

- -Contact - People with cellular contact are most likely to subscribe to a term deposit.

- -Month - May, June, July and August are the best times to get customers to suscribe.
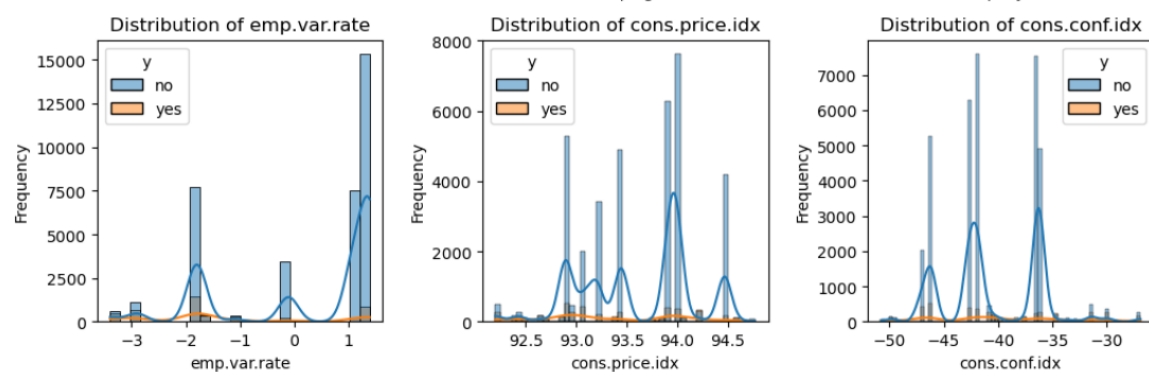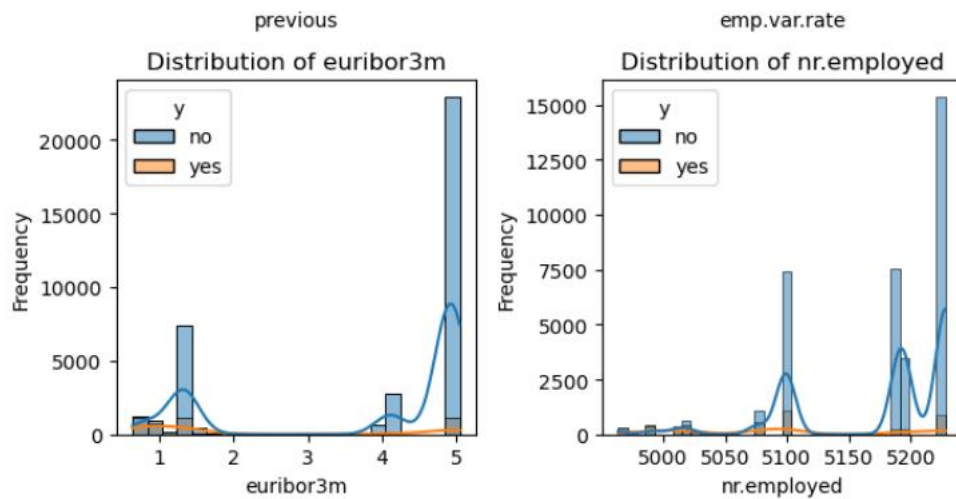
Histograms

Understanding Age Distribution:. This information is valuable to get marketing strategies to different age groups.

Campaign Contact Frequency: Creating histograms for "campaign" can help assess how frequently clients were contacted during the campaign.

Previous Campaign Contact: Histograms for "previous" can reveal the distribution of the number of contacts made with clients before the current campaign.
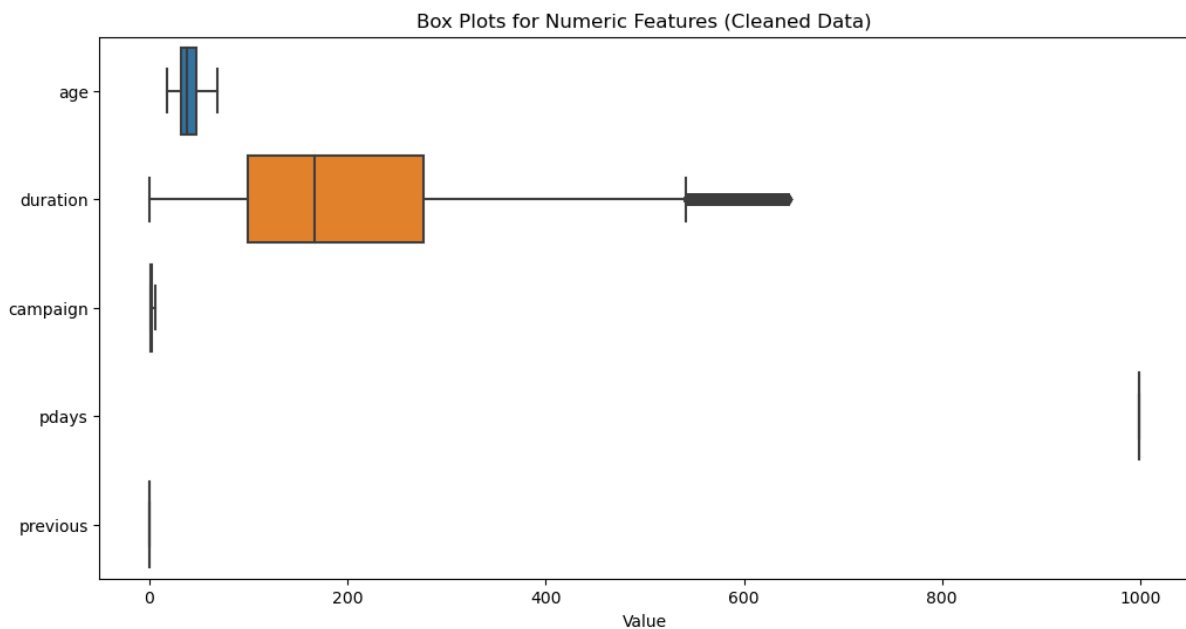
Economic Indicators: Histograms for economic indicators like "euribor3m" and "emp.var.rate" can provide insights into the distribution of these indicators, which may impact client decisions regarding term deposits.
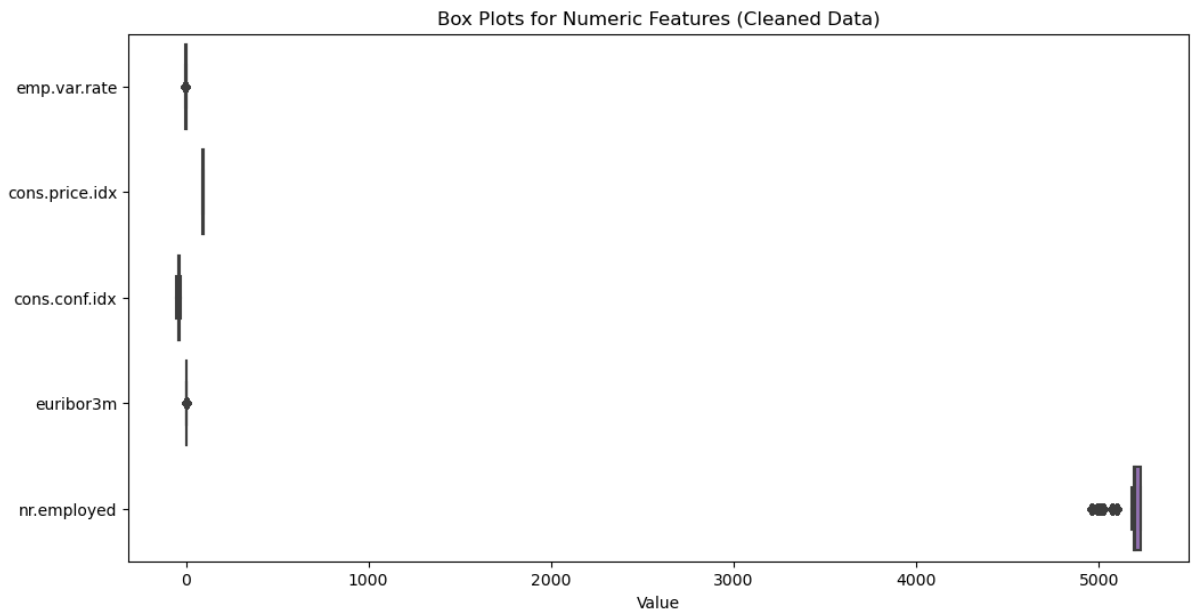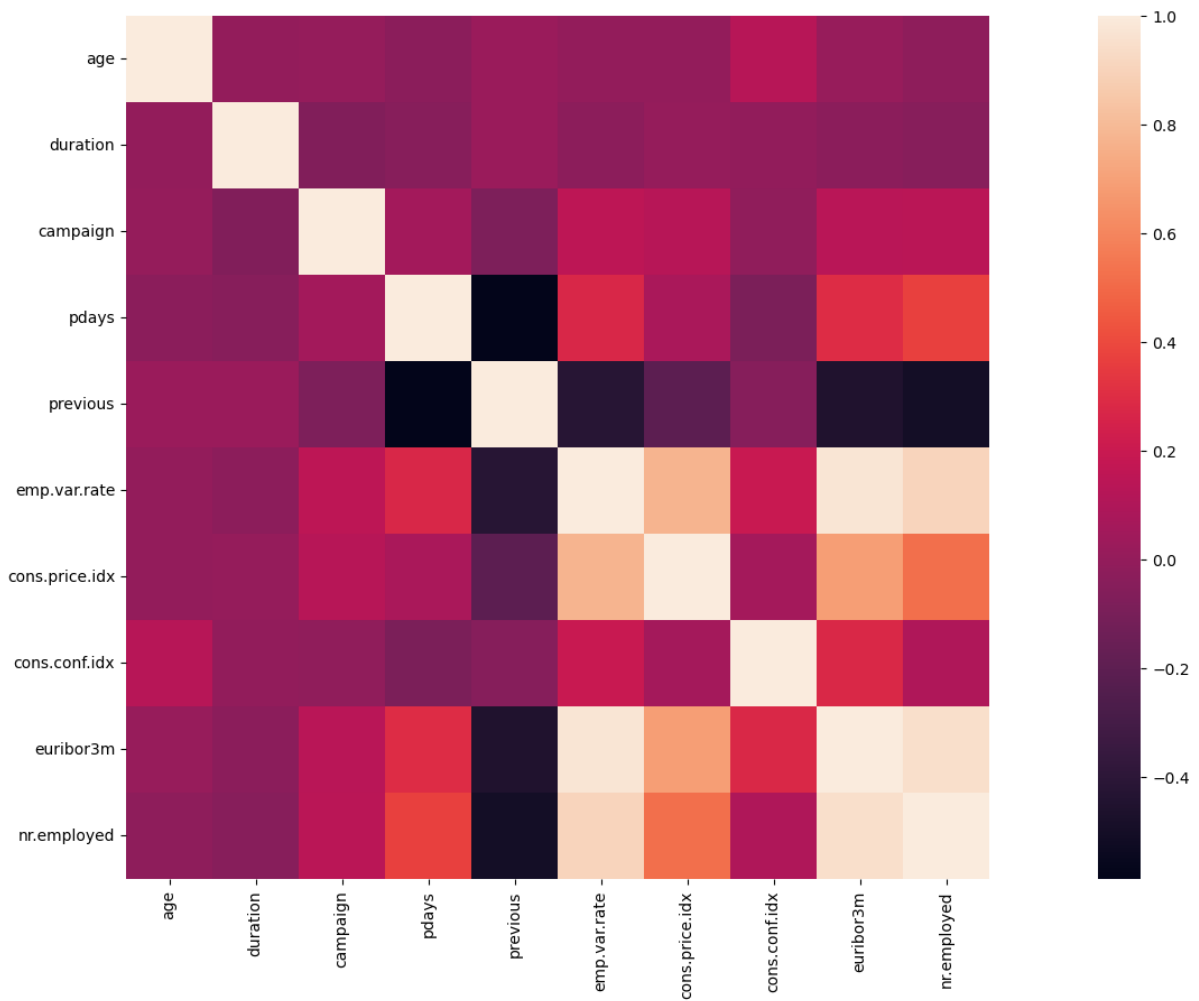
previous   emp.var.rate

Box Plots :

-
- From this we can indetify the outliers and thus remove it to improve the accurate results
- From this we can see that duration has many outliers and should be removed.

Box Plots for Numeric Features (Cleaned Data)

Correlation-Matrix

Heat Map Method to determine correlation between the continuous variables.

There was a strong positive relationship between emp.var.rateand euribor3m with correlation 97%.

The emp.var.rate attribute also had a strong positive correlation with nr.employed at 0.91 and with cons.price.idx at 0.78

Scatter Plots

By using scatter plots we can analyse the trends