Visualization Report - Bank Marketing Dataset

Debdutta Paul - 2981259

MAY 16, 2019

Part 1 – Exploratory Data Analysis

INTRODUCTION:

The problem domain: In cutting age, competition in banking domain marketing plays an important role. Marketing is planned execution of result, which gathered from analysis of long-term customer behavior, it is more important to know that which type of customer need to focus for successful outcome.

Title: Bank Marketing Data Set

The source of the data: https://archive.ics.uci.edu/ml/datasets/Bank+Marketing

The agencies working with the data: Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit.

The intended use of the data: Telemarketing is one the effective ways of marketing in banking domain and affectively resulting in good customer relationship. In this research, we are going to visualize the data regarding the term deposits based on our graphs we designed. Our predictions will lead to the success rate of taking term deposit on basis of 41,188 records present in dataset. Telemarking is the mode of marketing strategy used to contact customers for this dataset. There are about 21 variables, which include set of information of about 41,188 customers, which has been considered for visualization.

Attribute Information: There are 20 input variables and 1 output variable.

bank client data:

1. Age (numeric) – Age of clients being contacted.
2. Job : Type of job of the clients when being contacted (categorical: 'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown')
3. Marital : Marital status of the clients (categorical: 'divorced', 'married', 'single', 'unknown'; note: 'divorced' means divorced or widowed)
4. Education: The mentioned values are Clients educational background at the time of call. (categorical:'basic.4y','basic.6y','basic.9y','high.school','illiterate','professional.course',' university.degree','unknown'): default: has credit in default? (categorical: 'no', 'yes', 'unknown')
5. Housing: Does the client has housing loan? (categorical: 'no', 'yes', 'unknown')
6. Loan: Does the client has personal loan? (categorical: 'no', 'yes', 'unknown')

Note: All the above-mentioned attributes are related to the bank's customers.

related with the last contact of the current campaign:

7. Contact: Contact communication type (categorical: 'cellular', 'telephone')

8. Month: Last contact month of the year with the Client (categorical: ‘jan’, ‘feb’, ‘mar’, ..., ‘nov’, ‘dec’)
9. Day_of_week: Last contact day of week with client (categorical: ‘mon’,’tue’,’wed’,’thu’,’fri’)
10. 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.

Note: All the above-mentioned(7-10) attributes are the last contact info where bank has contacted the clients.

other attributes:

11. Campaign: Number of contacts performed during this campaign and for this client (numeric, includes last contact)
12. PassedDays: Number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means clients were not previously contacted)
13. Previous: Number of contacts performed before this campaign and for this client (numeric)
14. Poutcome: Outcome of the previous marketing campaign (categorical: ‘failure’, ‘nonexistent’, ‘success’)

Note: All the above-mentioned(11-14) attributes are the other additional info where bank has tried to organize campaign or they have tried to contact the clients for term deposits.

social and economic context attributes

15. Emp.var.rate: Employment variation rate - quarterly indicator (numeric)
16. Cons.price.idx: Consumer price index - monthly indicator (numeric)
17. Cons.conf.idx: Consumer confidence index - monthly indicator (numeric)
18. Euribor3m: Euro Interbank Offered Rate (Euribor) 3-month rate - daily indicator. Based on the Average Interest Rates at which a large panel of European banks borrow funds from one another that mature after 3 months. (numeric)
19. Nr.employed: Number of employees - quarterly indicator means number of employed persons for a quarter. (numeric)

Note: All the above-mentioned(15-19) attributes are the social and economic attributes which gives the information about monthly or quarterly indicators and different interest rates.

Output variable (desired target):

20. Output - has the client subscribed a term deposit? (binary: ‘yes’,’no’)

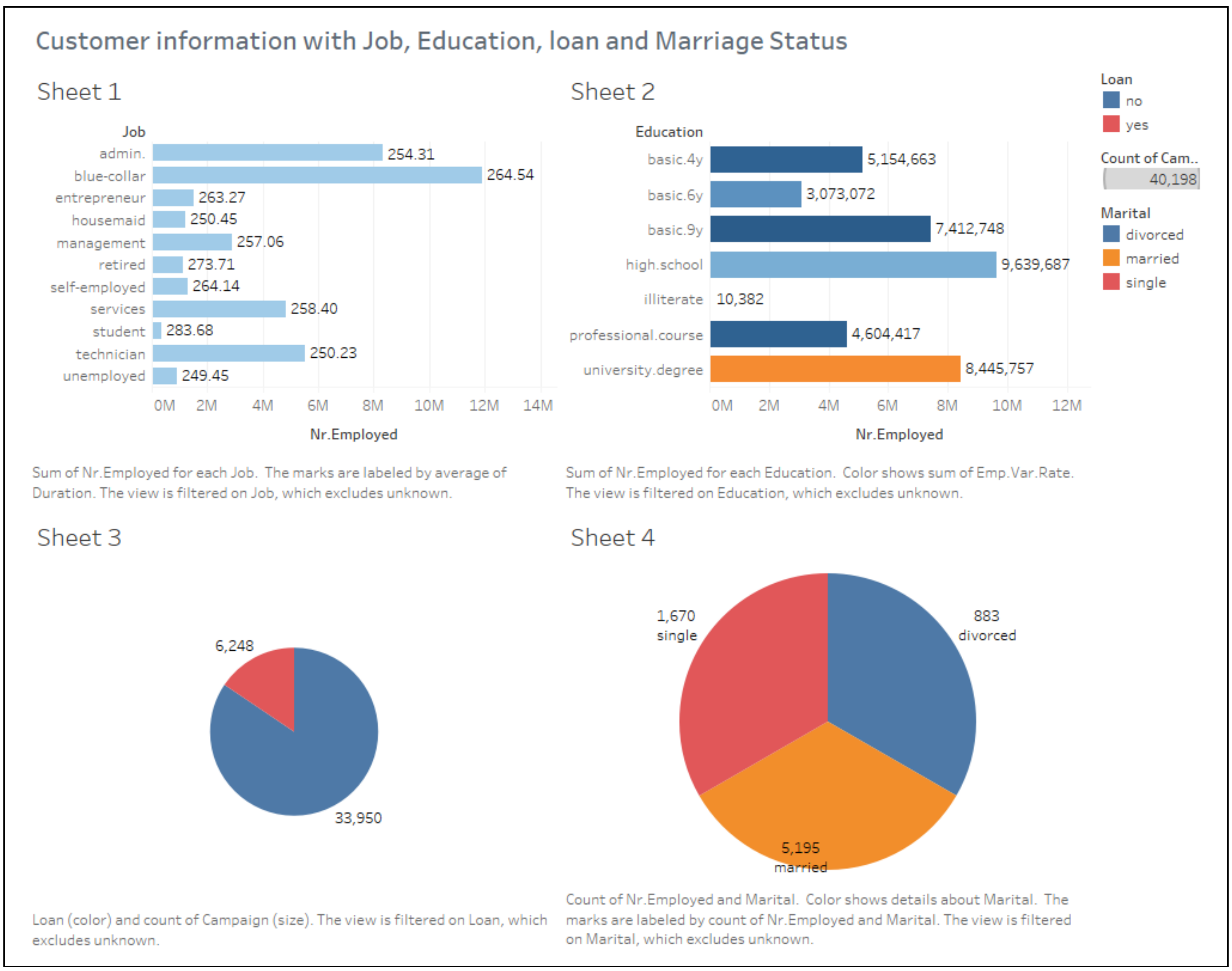
Objective:

This dataset is taken from a Portuguese Bank which is conducted to direct marketing campaigns to promote term deposits to their customers. The common problem could be that the bank keeps re-calling the ‘‘wrong’’ customers (the ones don’t need term deposits) and it may cause high labor costs and be susceptible of harming customer relationship. So, our business goal is to improve marketing effectiveness by targeting the right customers.

Problem Statement:

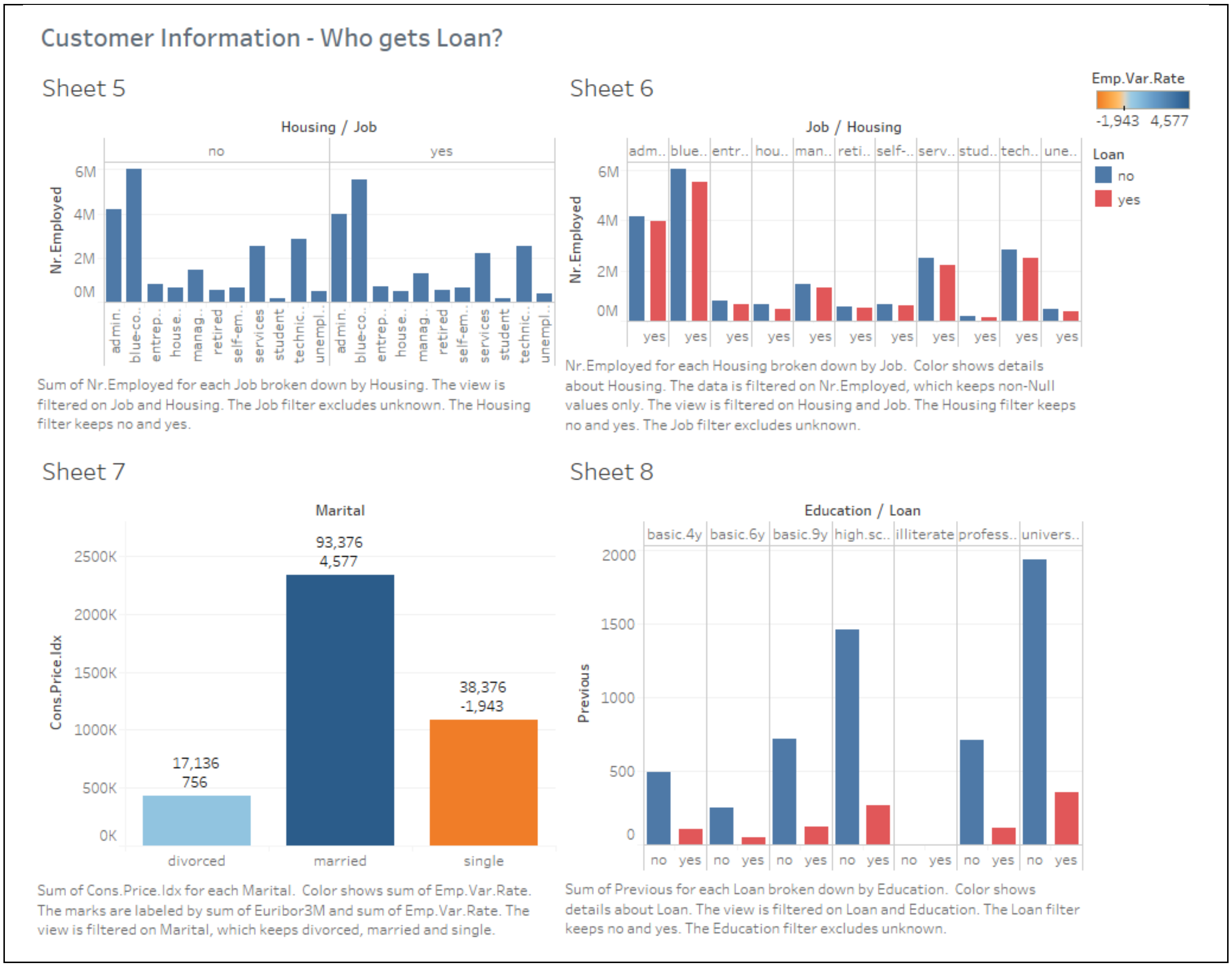
In this project, the stakeholders are the bank marketing team, bank employees, and customers. The bank would benefit from this solution where lower marketing costs and less probability of wrong marketing targets can be achieved without the risk harming customer relationship. The customers of the bank will receive more precise solutions to their needs. However, this solution may have problem where new customers have no previous marketing records and thus reduce the size of available features. Also determining which category of customer (single, marriage or divorced) gets the loan easily.

Dashboard 1:



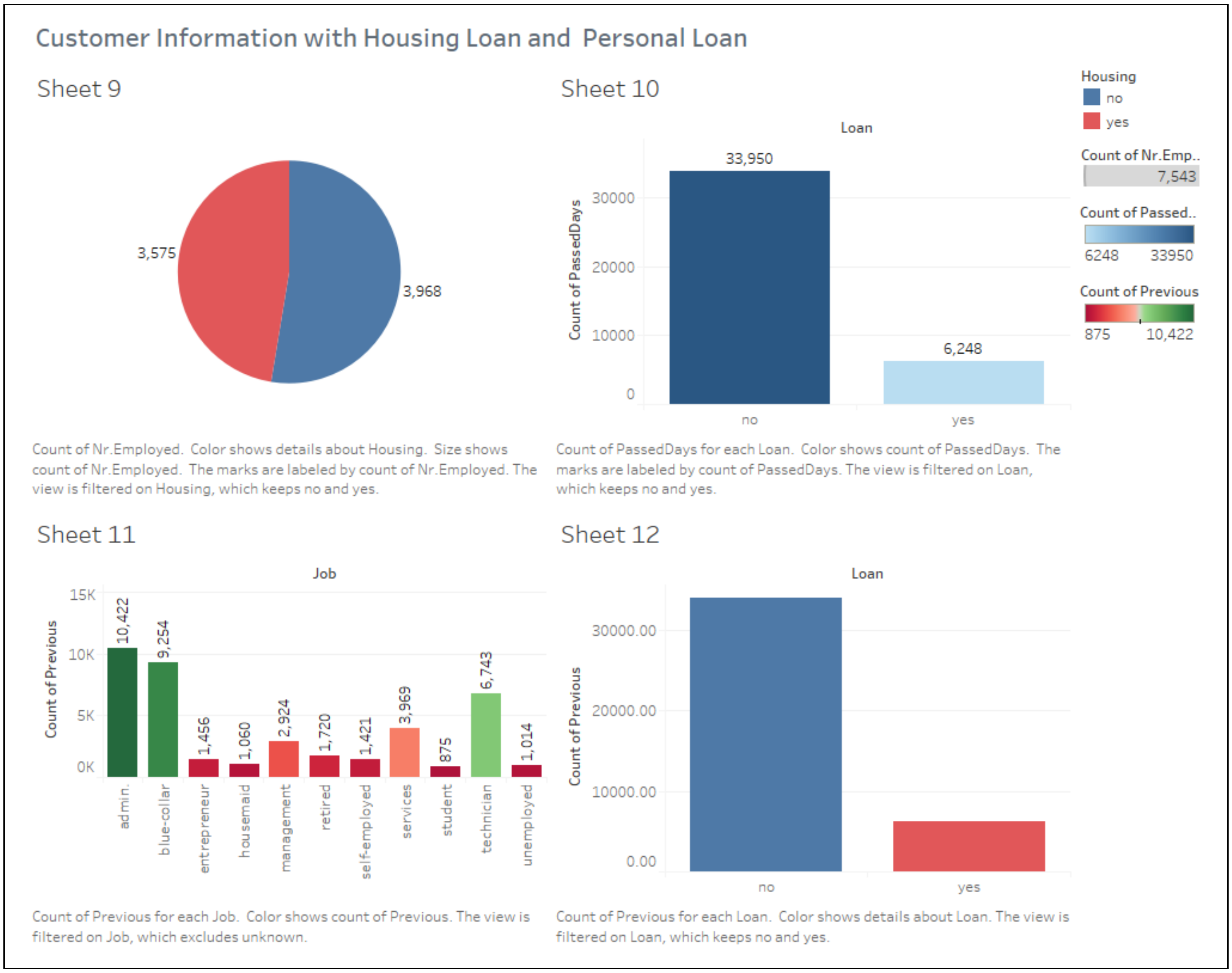
The overall dashboard tells us the count of campaigning done for personal loan and how many numbers of employees are educated in each category with the amount of people holding a desired job such as admin, technician, student etc. Also, it shows the marital status of the person when they are being contacted for personal loan.

Dashboard 2:



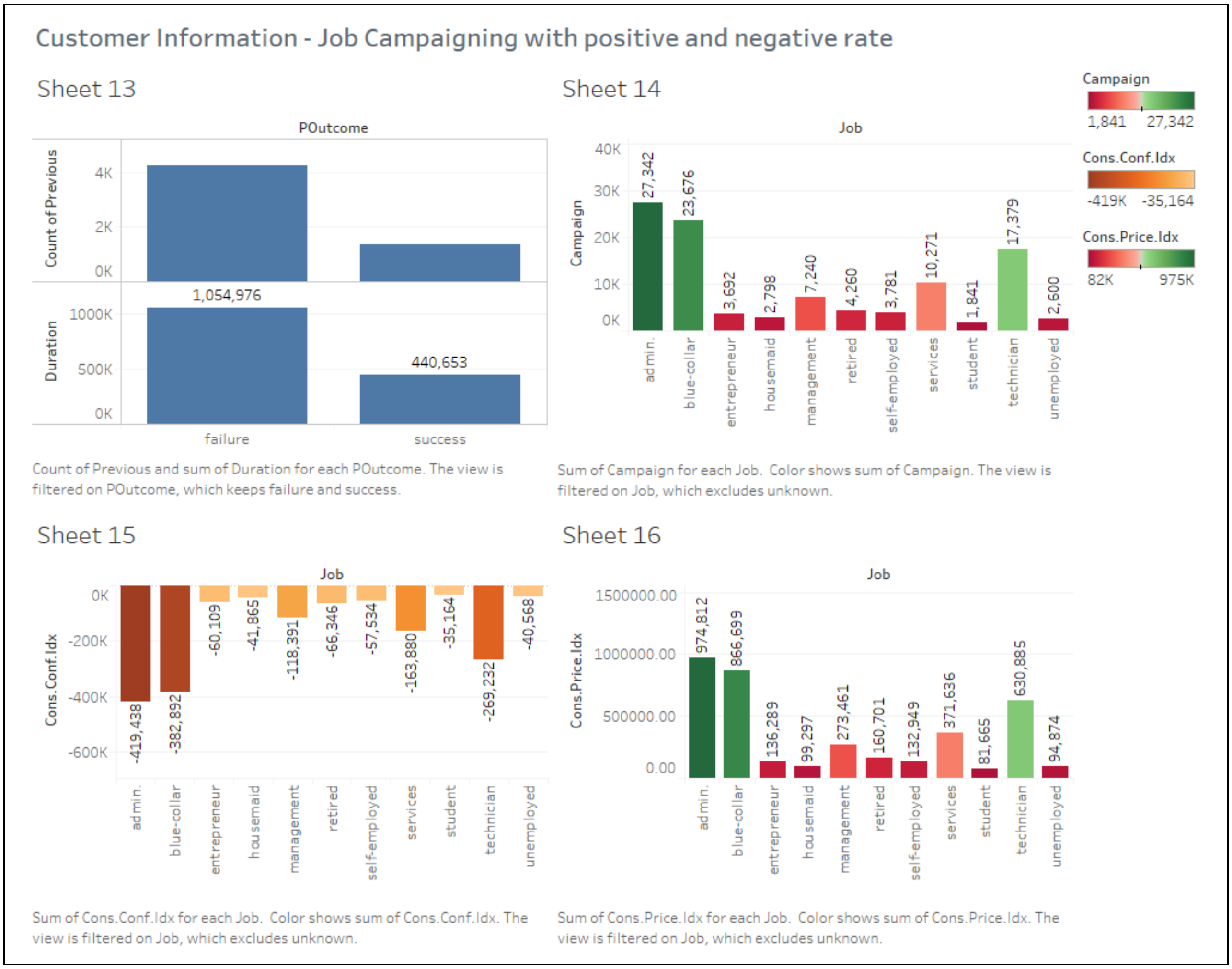
The above dashboard defines the relationship of different variables of different customers and tells us which customer gets the loan. The relationship shows number of employees with both housing loan and job. Also, it shows the sum of consumer price index with employee variable rate which is filtered based on marital status. Finally, there is a breakdown of personal loan based on education. Here we can observe that marital people get highest personal loan as the employee variable rate is high.

Dashboard 3:



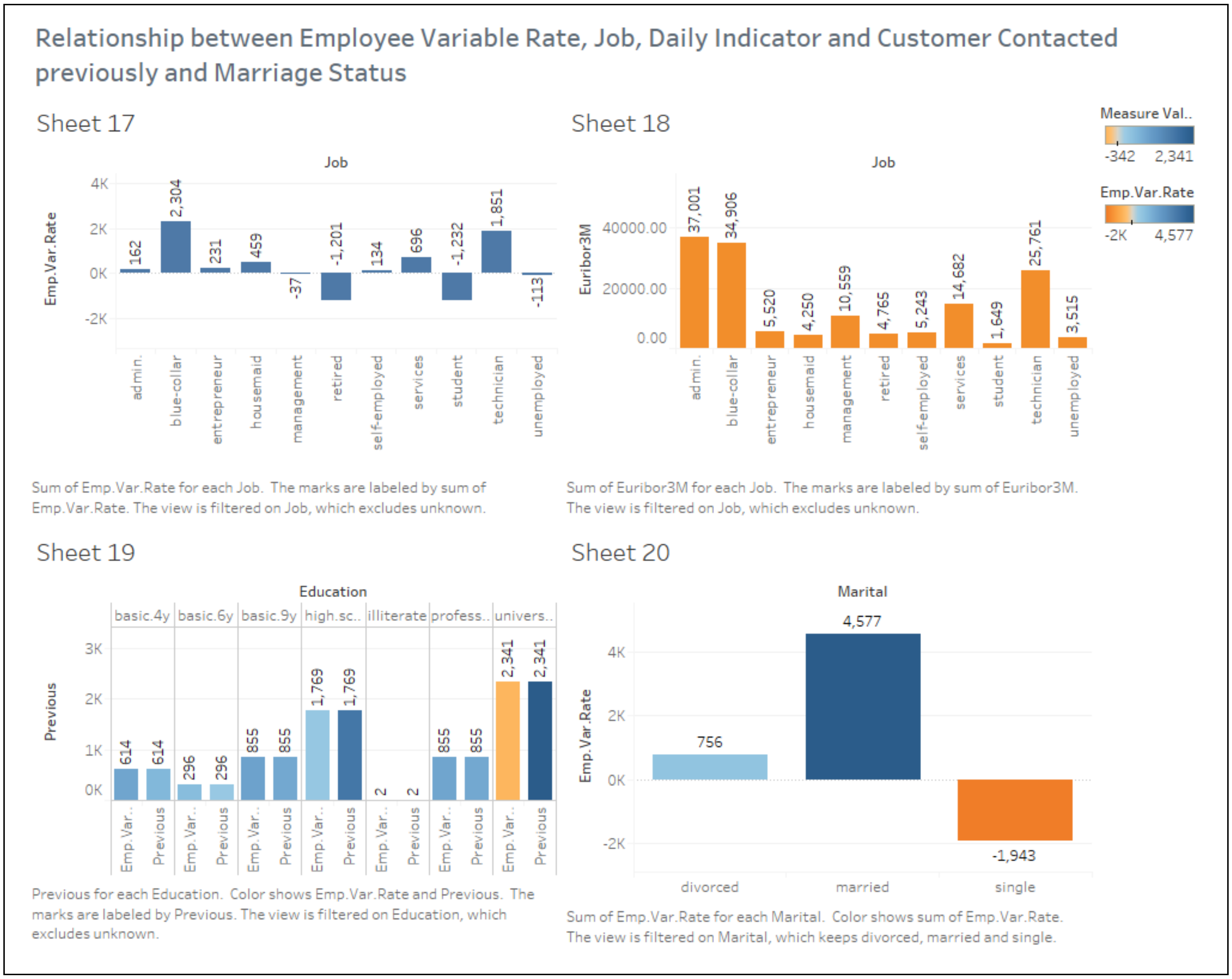
The dashboard describes the customer information where they have both housing loan and job to get personal loan. The pie-chart in the above dashboard describes the number of employees who has housing loan. The next bar-chart describes the count of the number of days that passed by after the client was last contacted from a previous campaign. Also, there is a graph that shows the count of number of clients contacted before the previous campaign. Finally, the last graph shows the count of previously contacted customers and their loan status. Most of the previously contacted customers do not have loan.

Dashboard 4:



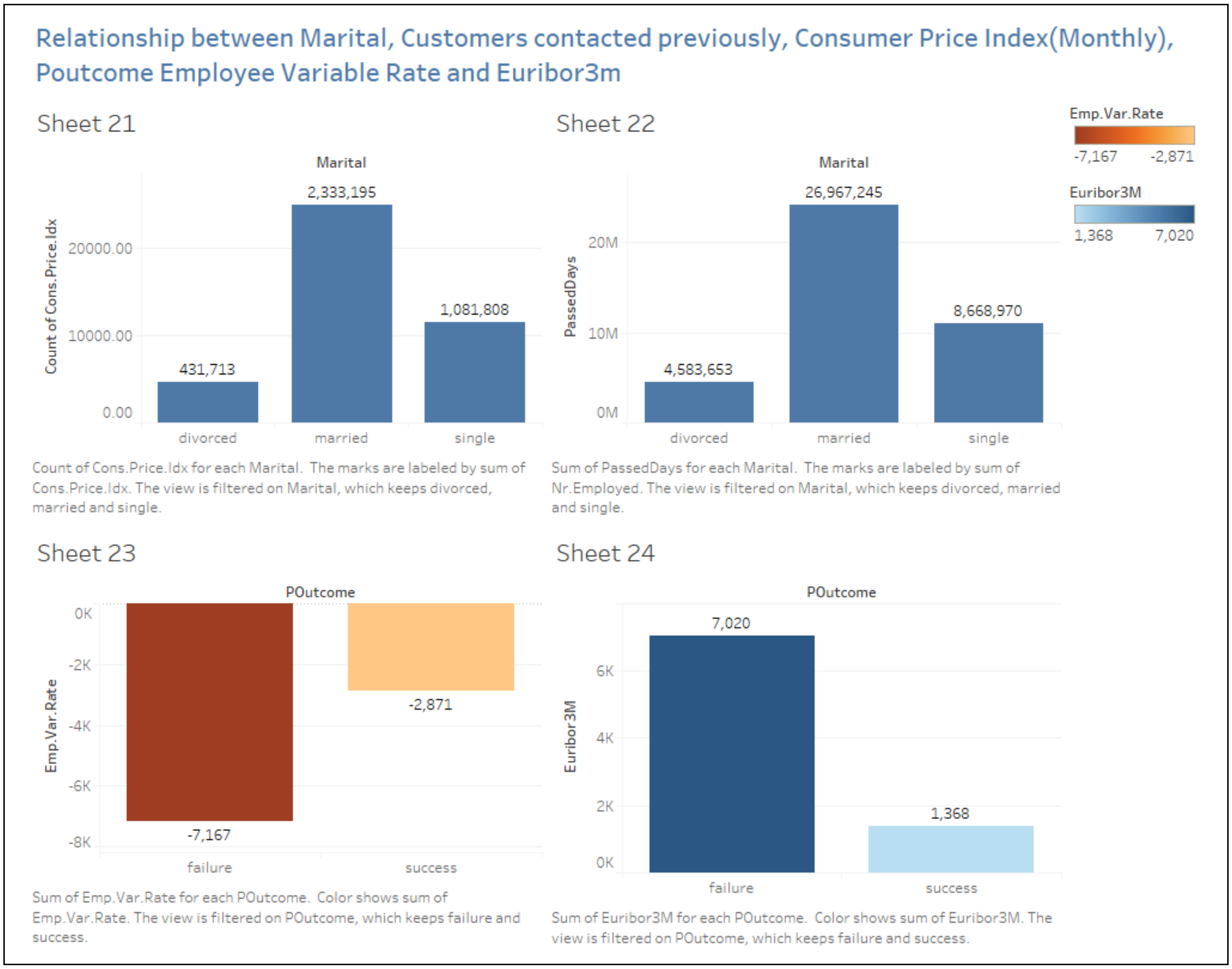
The above dashboard defines count of previous contacted customers and the sum of their duration based on the success and failure rate of previous campaign. Different color codes define the sum of campaign done for each job which results the sum of consumer confidence index as negative and consumer price index as positive.

Dashboard 5:



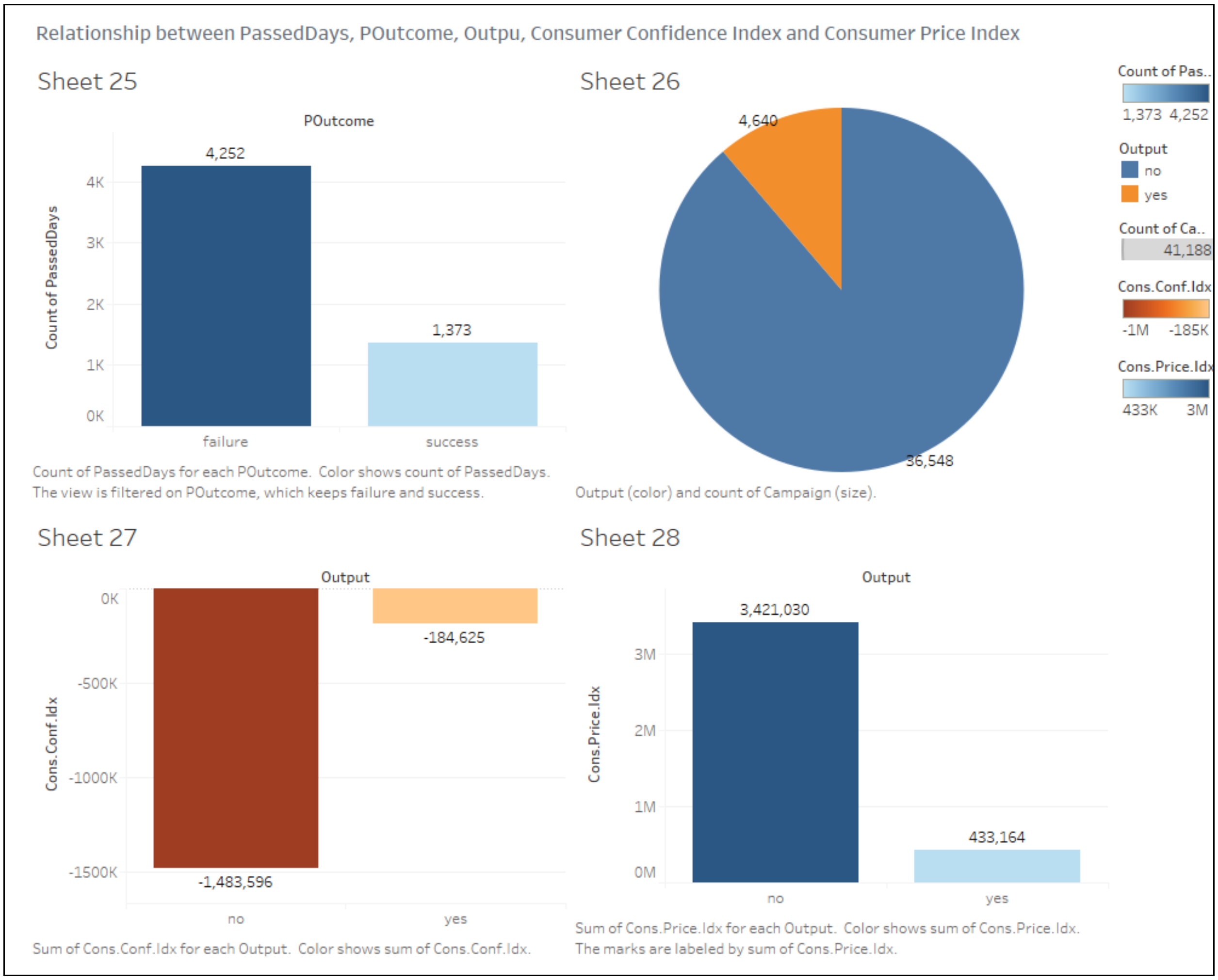
The dashboard above explains the employee variable rate and euribor3m with each job. Also, it shows the sum of total number of contacts performed before a client with respect to education. There is also a graph describing the total value of employee variable rate where it shows the marital status of the employees. Overall it says the married customers with maximum employee variable rate gets the loan.

Dashboard 6:



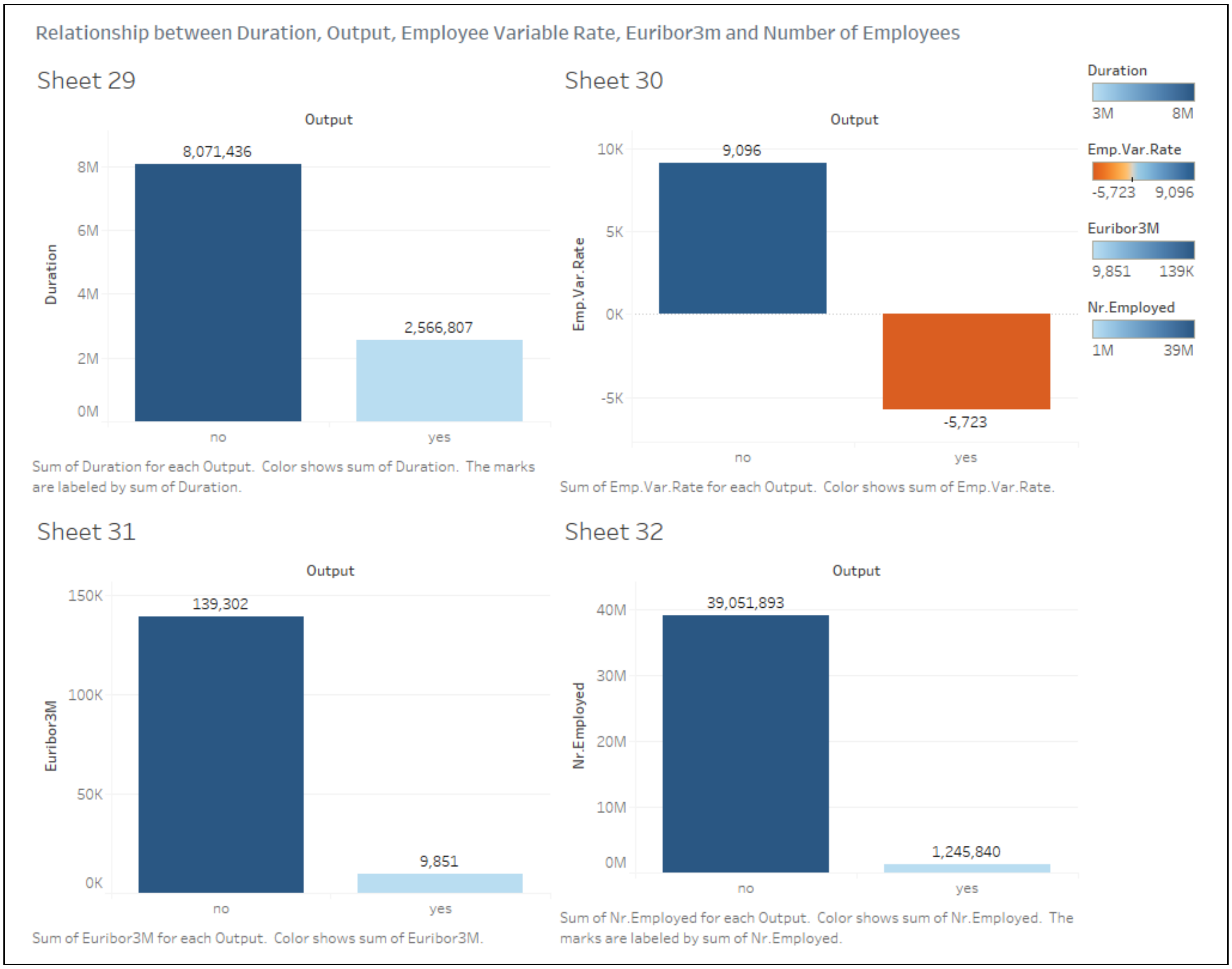
The dashboard explains the count of consumer price index and sum of days passed for each marital based on number of employees. The sum of employee variable rate defines the higher success rate over rate being negative values. Also, it has the interest rate defined with each outcome where failure rate takes over success rate.

Dashboard 7:



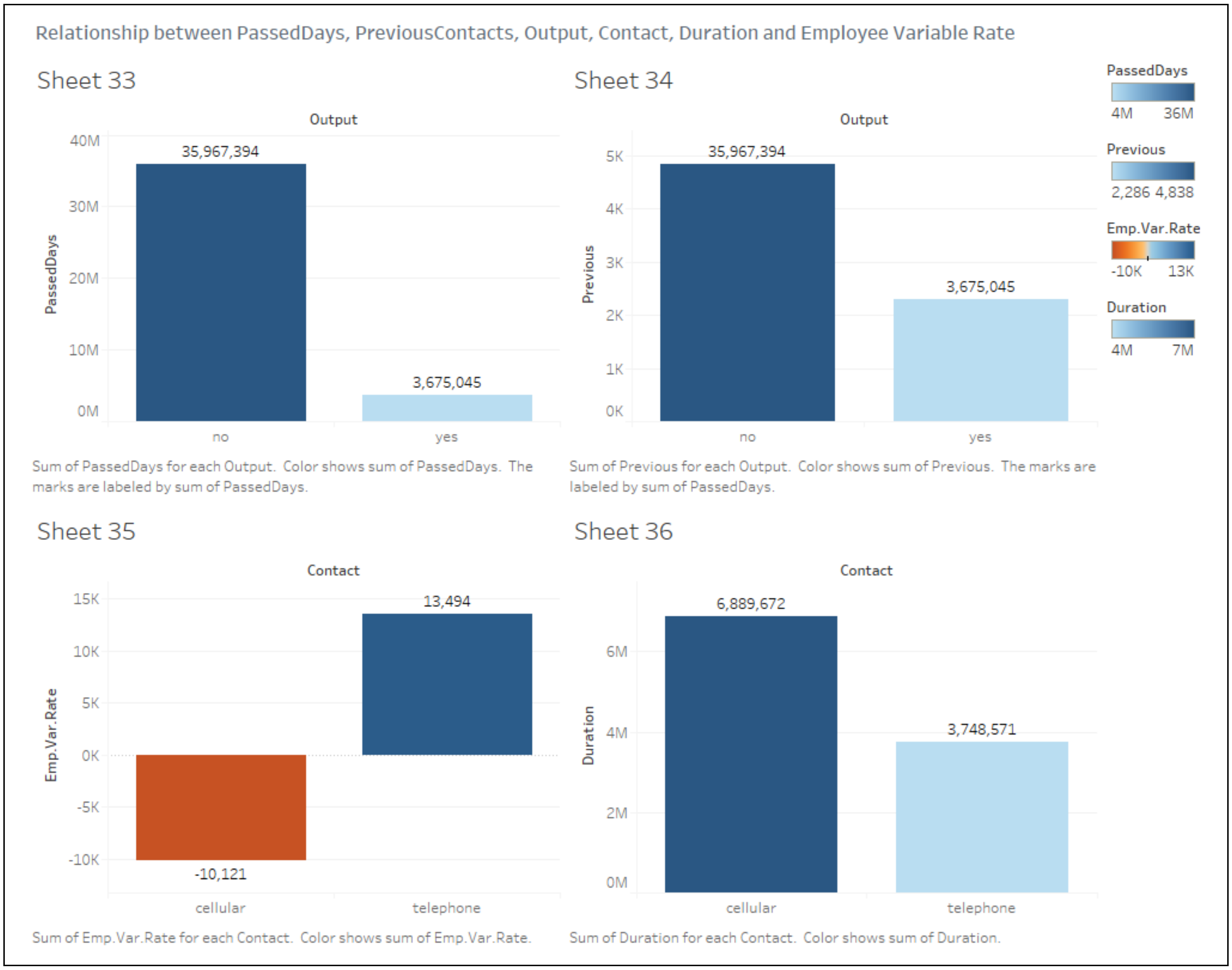
The dashboard explains the number count of days passed for each outcome where success rate is very low. As we see in the pie-chart, majority of the client is not interested in taking loan and has not subscribed for term deposits, based on the campaigning done. There is a consumer confidence index with negative values where the values of the output are better in success rate, but the success output value of consumer price index is less.

Dashboard 8:



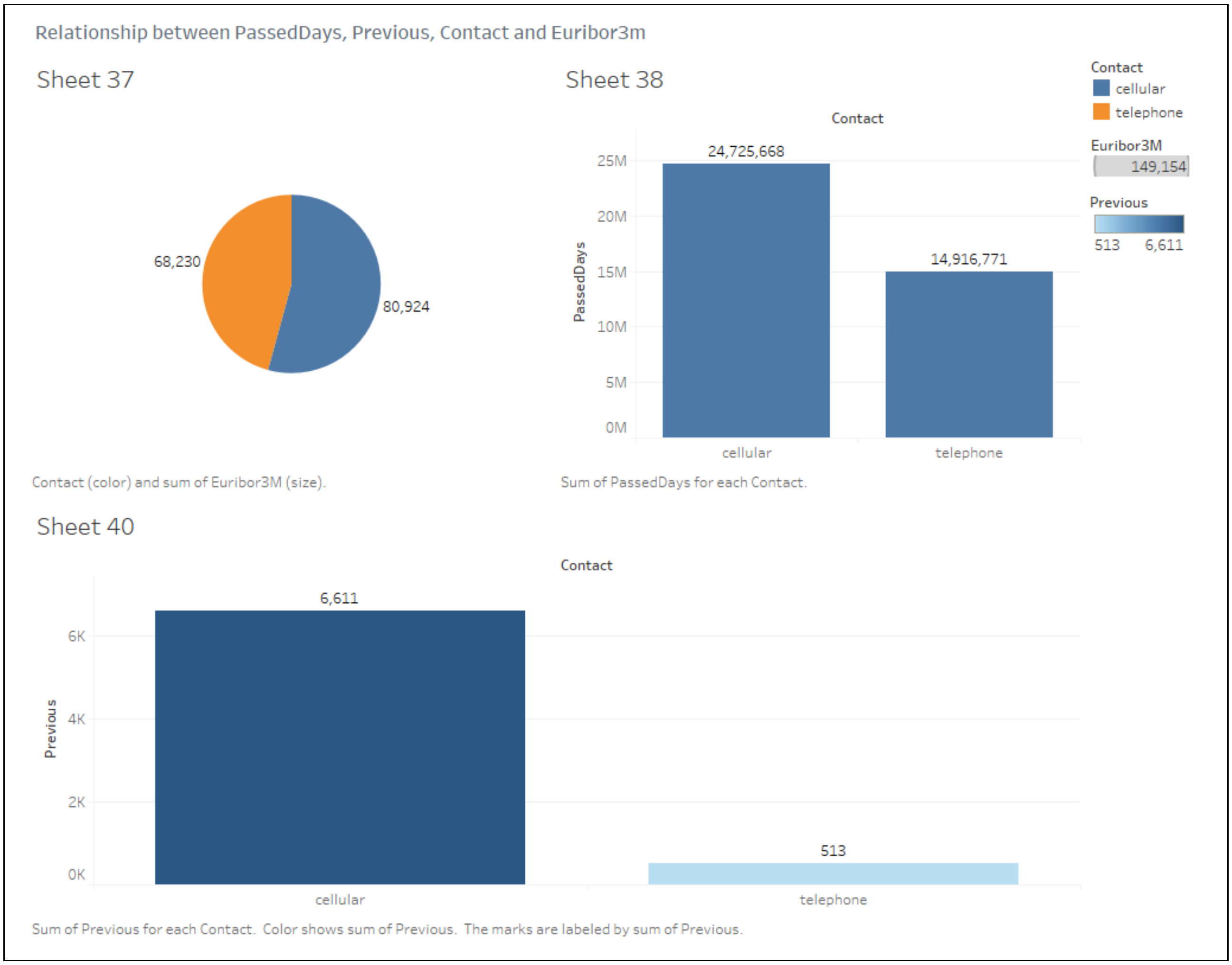
The dashboard defines four sheets where first sheet shows us the total duration spent on call to the clients who are interested in taking loans and are subscribed to term deposits. The other sheet shows the total sum of employee variable rate along with average interest rate and total number of employees, which results for each output.

Dashboard 9:



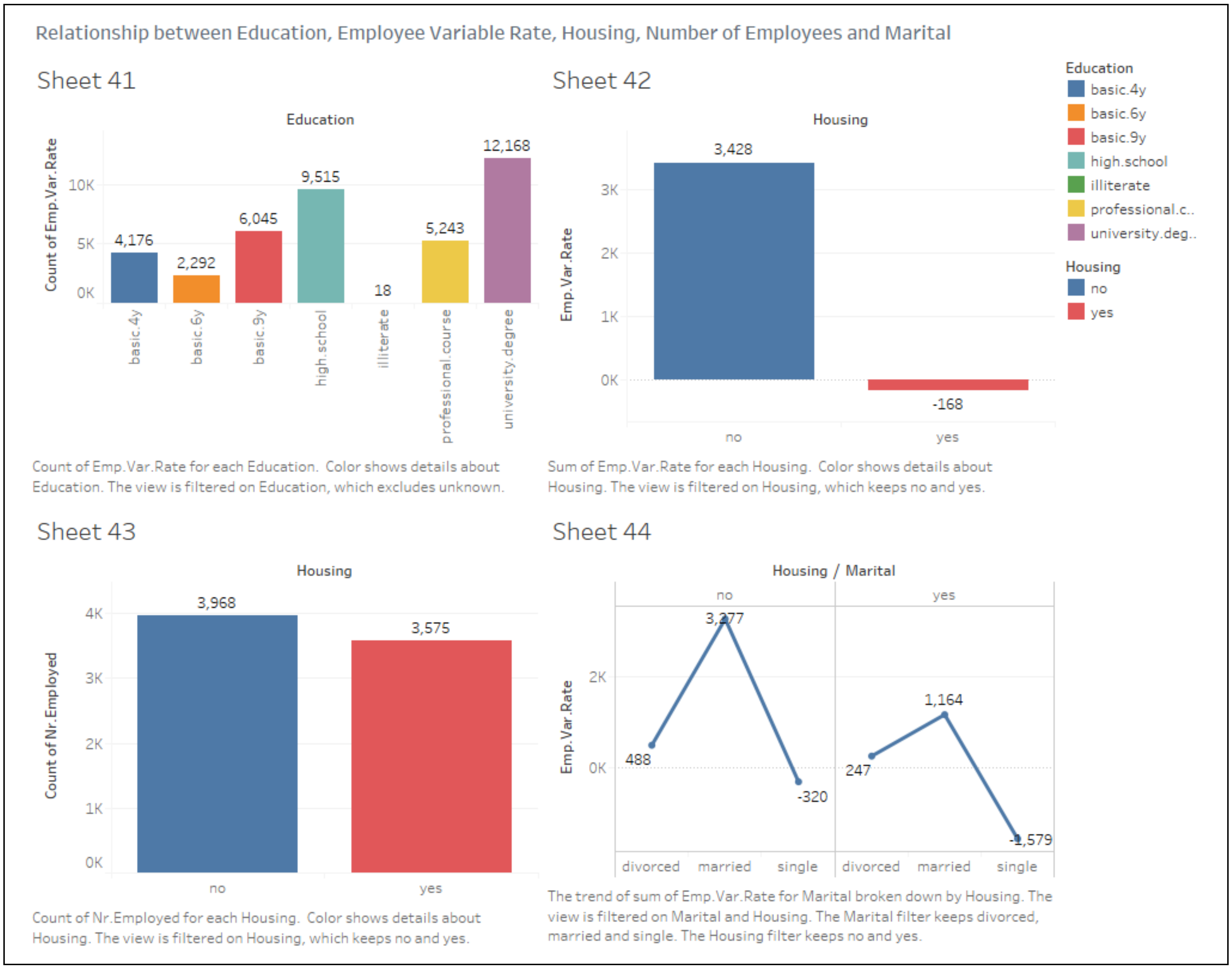
The dashboard has four sheets which defines the sum of total number of days passed by the client when contacted over call whether the client has subscribed in term deposit. Also, there is graph where the number of customers contacted previously for term deposits. The next sheet tells about the employee variable rate of each employees when contacted over cellular and telephone. The last sheet shows the time duration of each calls contacted by cellular and telephone.

Dashboard 10:



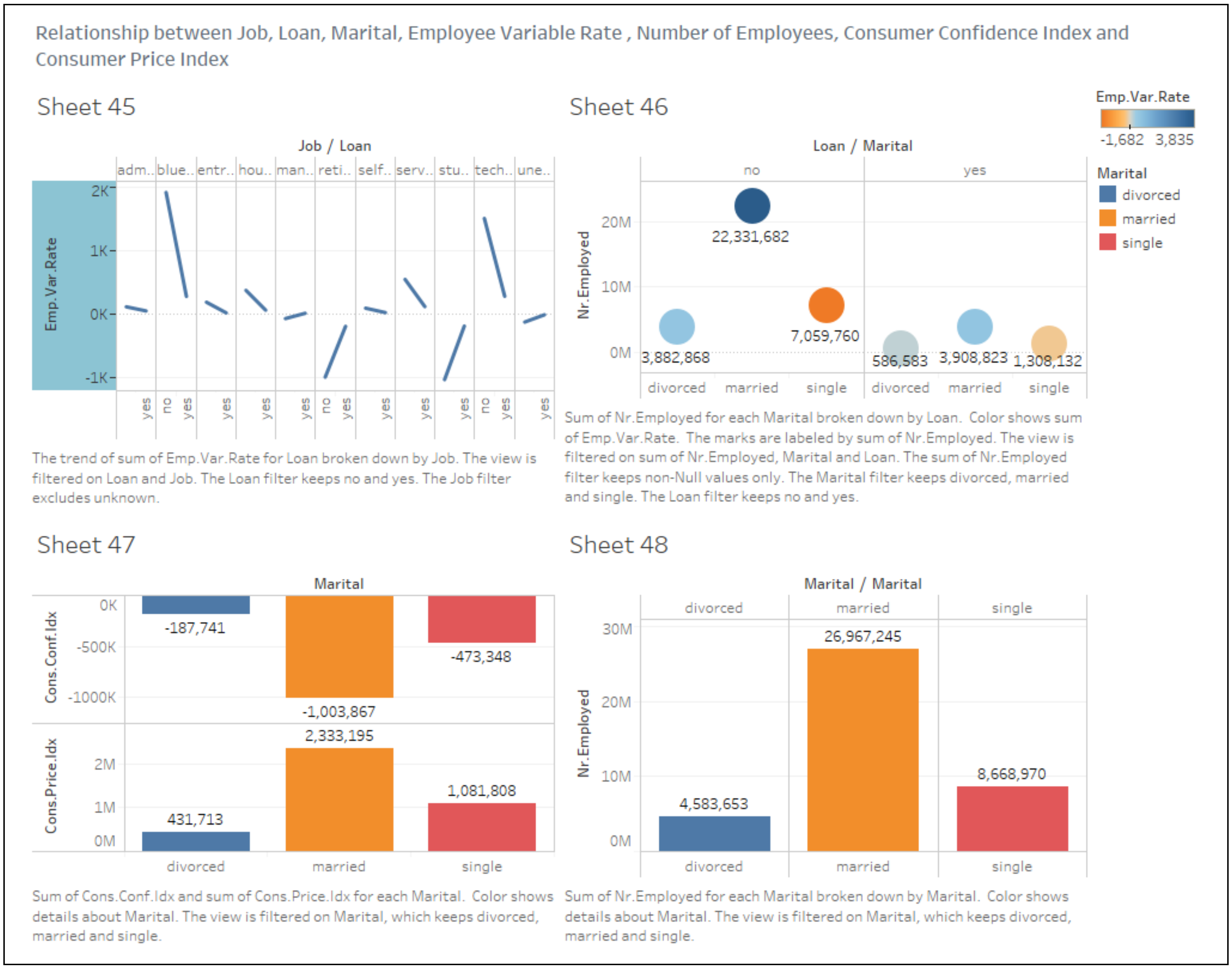
The above dashboard shows the sum of average interest rate, number of days passed after the client was contacted and sum of the total number of contacts done for the client over cellular and telephone.

Dashboard 11:



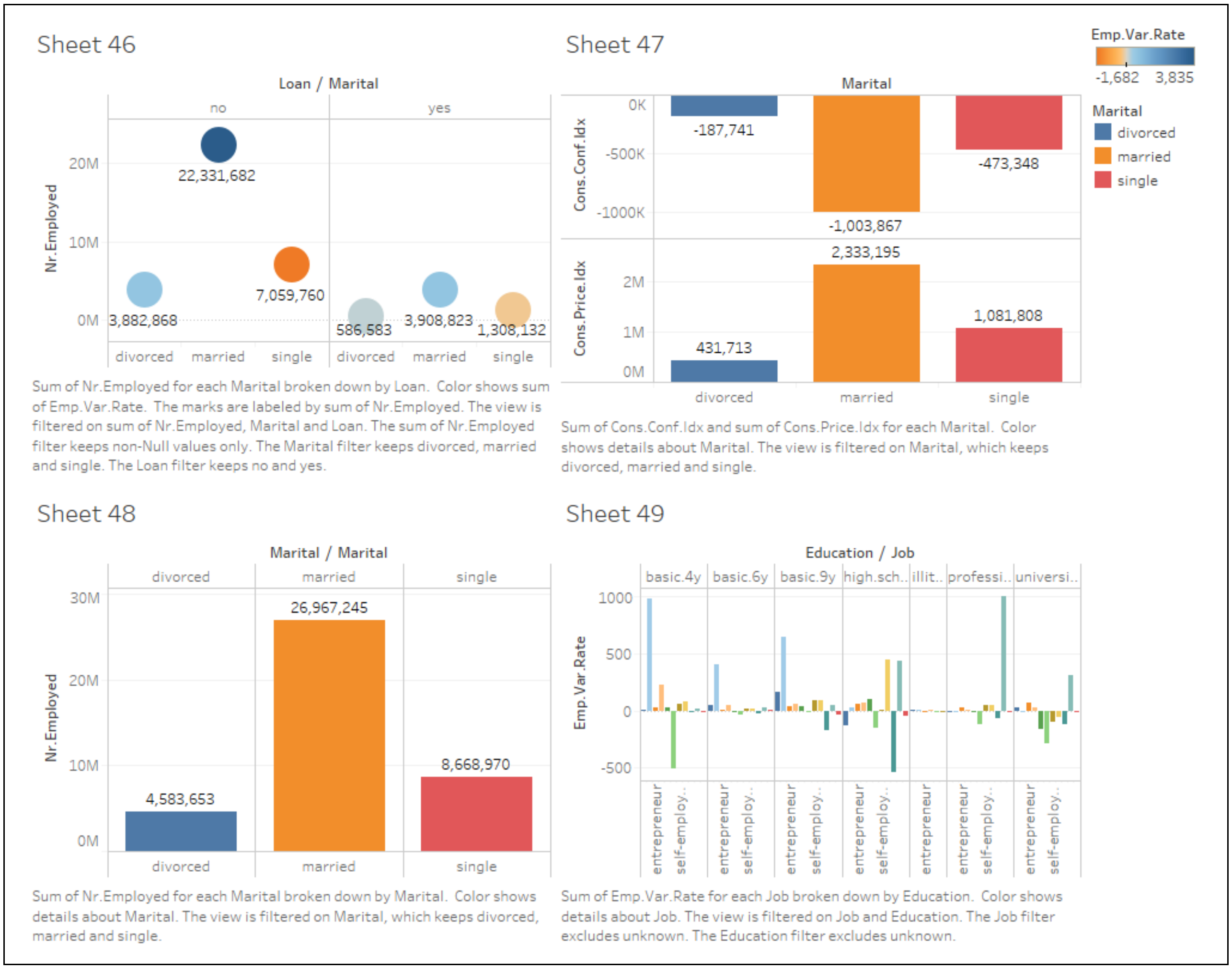
The above dashboard defines the relationship between different variables. The count of employee variable rate for total employees based on education. There is a value where number of employees displays whether they have housing loan or not. Also, employee variable rate is high whether the employees are married or not and they have housing loan or not.

Dashboard 12:



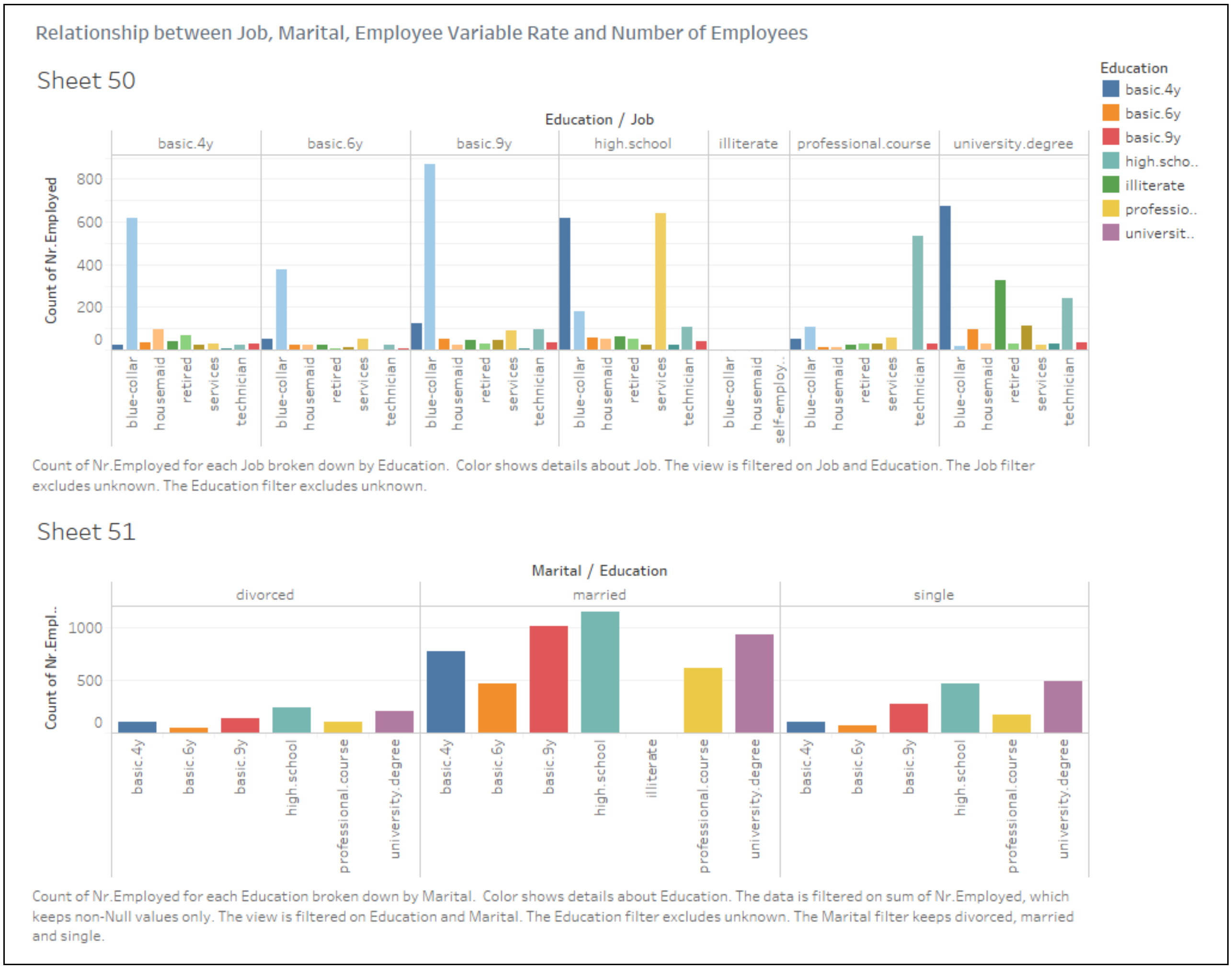
The above dashboard explains the trend of sum of employee variable rate for loans by per job. Also, it shows the loans initiated by married people are higher than compared to singles and divorced people. The sum of consumer confidence index has better result than married and single with being negative values. In other way, if we see the consumer price index, married people have better result than single and divorced. Overall, we see, the number of employees who are married, gets the loan easily.

Dashboard 13:



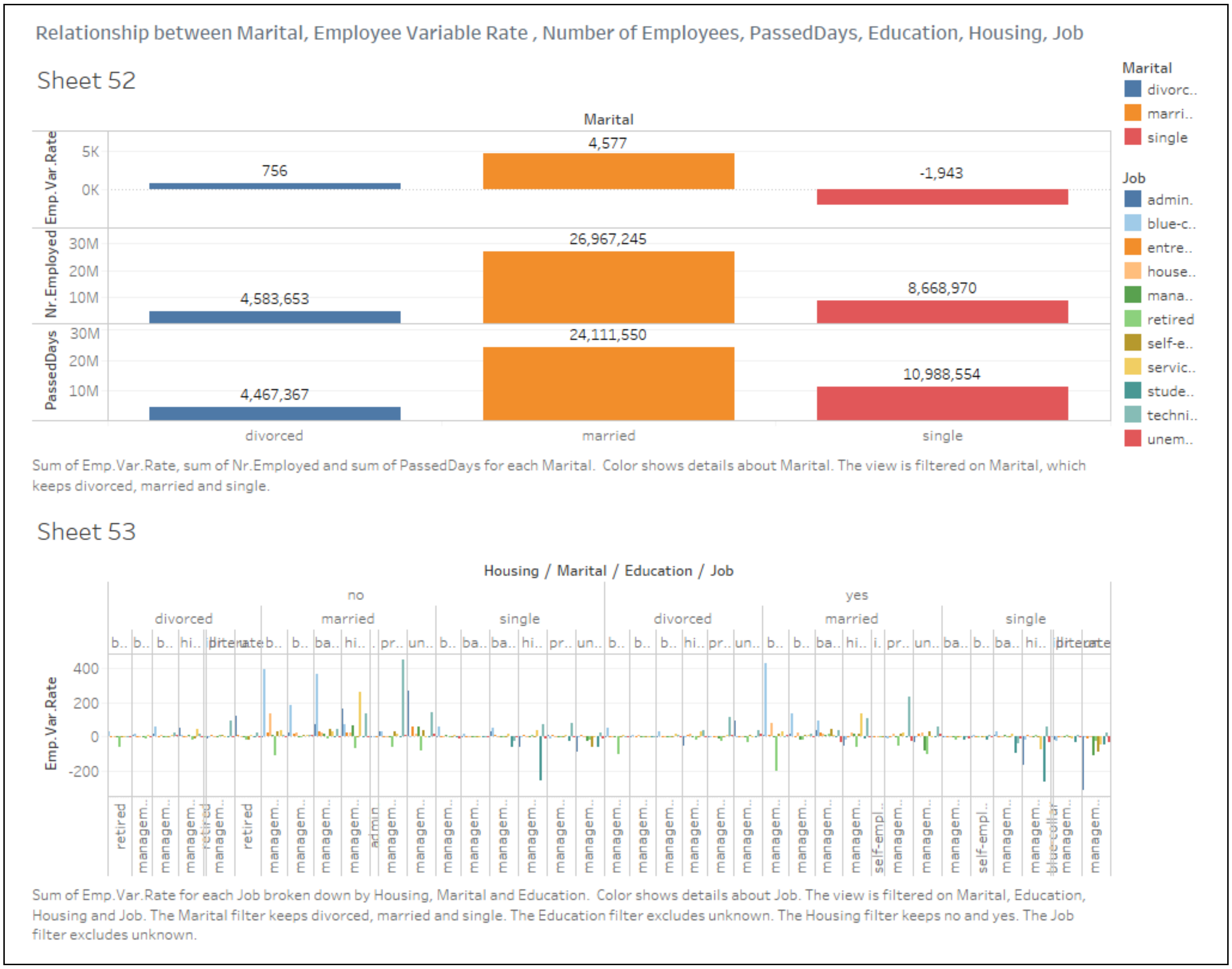
The above dashboards outline the sum of each number of employees with marital and loan. Also, it tells us about the sum of consumer confidence index where the value of divorced people is better than married and single. But in case of consumer price index, the value of married people is more than the values of single and divorced. The next sheet says about the number of employees with their marital status. There is a relationship between the sum of employee variable rate for each employee based on education and job.

Dashboard 14:



The above dashboard defines count of number of employees based on their education and job. Also, it shows the count of number of employees based on their education and marriage status.

Dashboard 15:



The above dashboard displays the sum of employee variable rate and the sum of days passed by after the client has last contacted from a previous campaign which is filtered based on marital. Also, it displays the employee variable of the employees with respect to Housing loan, Marital, Education and Job. The graph is filtered and is analyzed based on all mentioned variables.

Conclusion:

From all the above dashboard, it can be concluded that employees who are educated, married, employed and have a housing loan will get the subscription of term deposits and will get a loan easily. Also, it can be said that the married customers have high consumer confidence index and consumer price index than single and divorcee. Also, it can be concluded that failure rate is higher than success rate which is determined from the campaigns conducted by bank. Customers who spend more time speaking with bank employee via telephone seems to be interested in the subscription of term deposits.

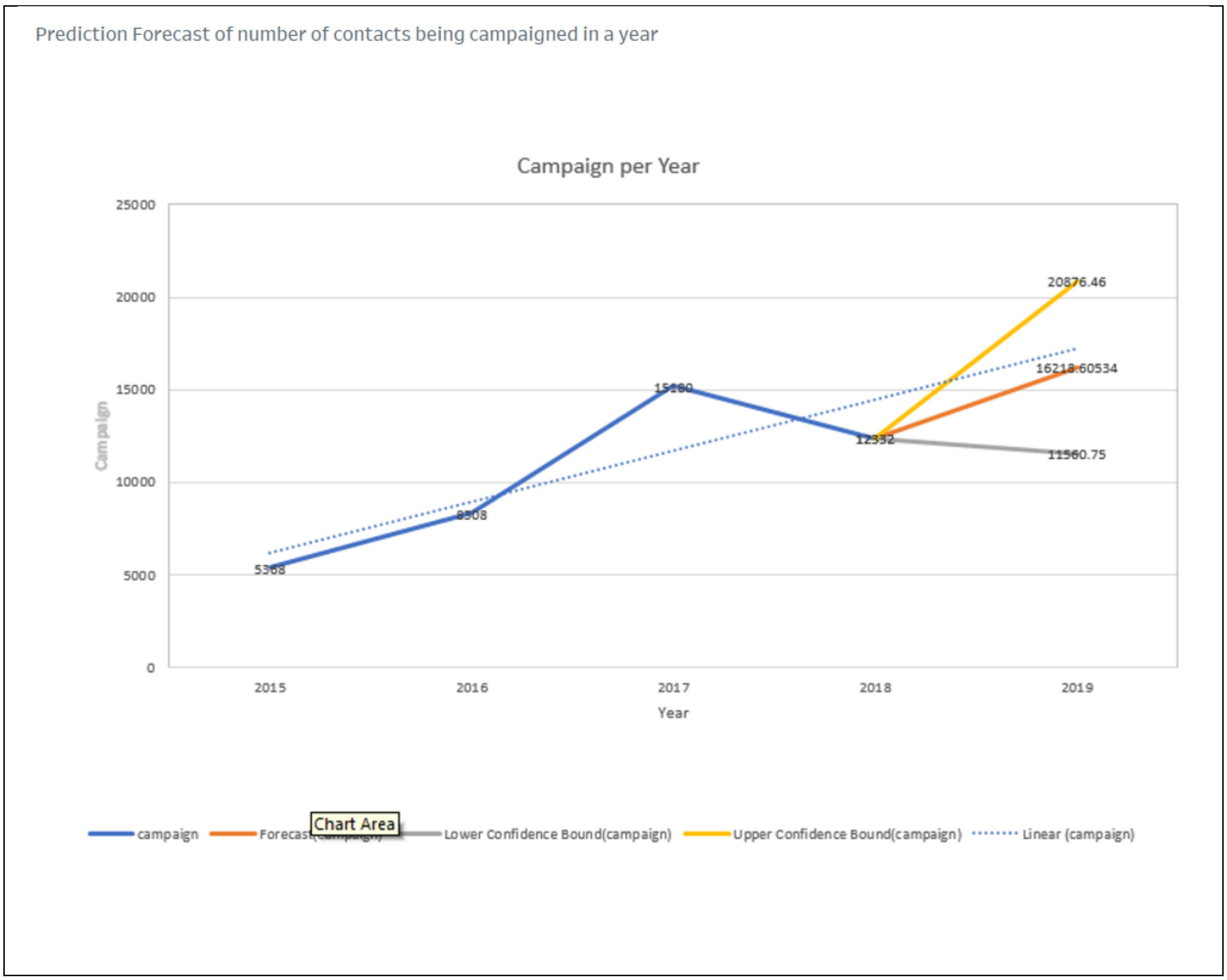
Part 2 – Predictive Analysis on Forecasting

Forecasting is the term to predict the future value. It is a statistical technique which helps in predicting future value.

In this project, for predictive forecasting, the technique used is exponential smoothing. This technique observes the older observations and captures the emerging trend of the data and explore them to the future. The result of forecast becomes a visualization to the user.

Below observations are done based on a time dimension and a measure field to create a forecast. Forecasting model is performed in excel sheet first and then it is pulled to tableau dashboard and finally a story has been made and a report is created.

Dashboard 16:



Predictive Forecasting:

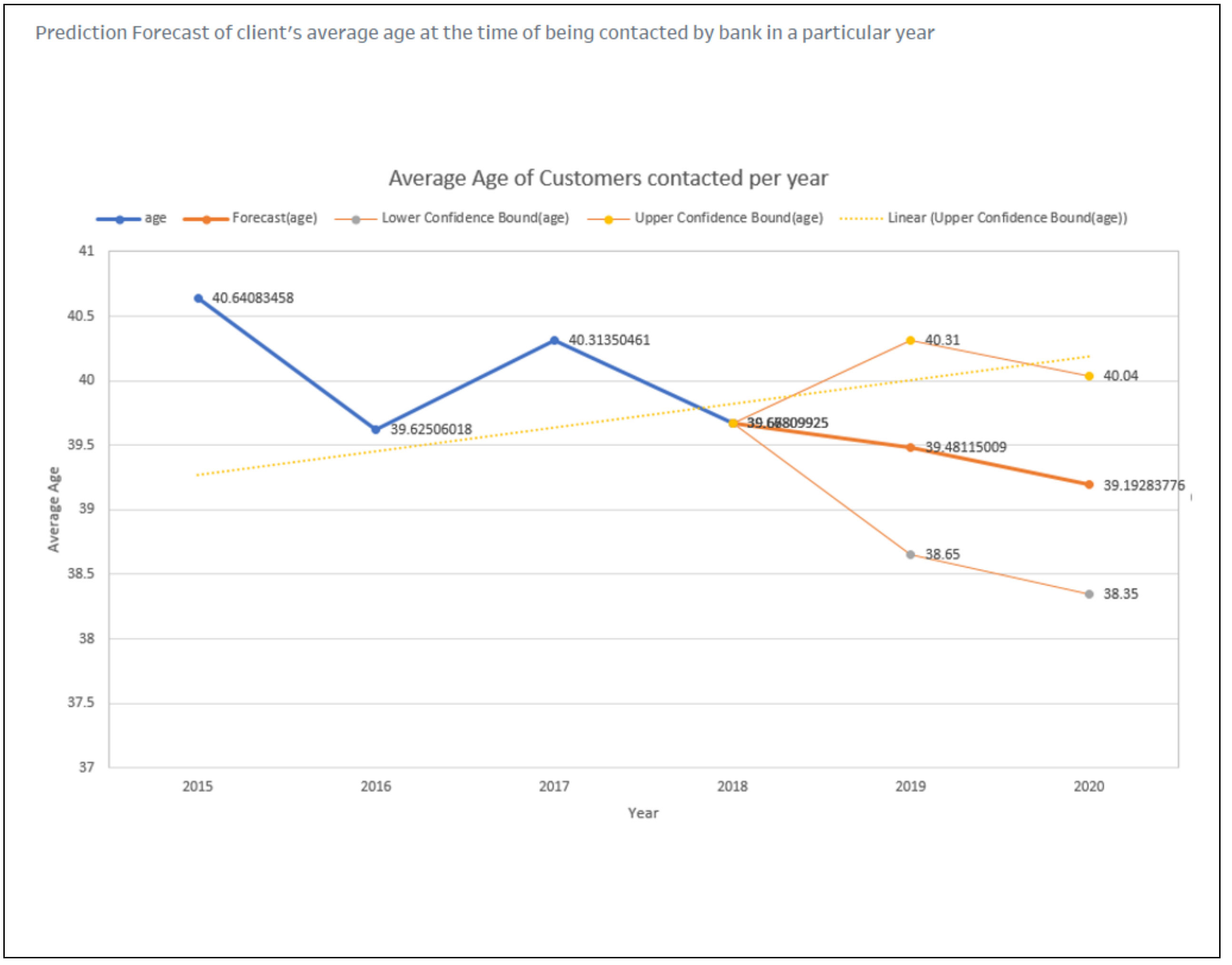
Time series:	Year
Measures:	Campaign
Forecast forward:	1 year (2018 – 2019)
Forecast based on:	2015 – 2018
Confidence Interval	95%
Seasonality	As the data is yearly, we cannot support seasonality option over here.

year	campaign	Forecast(campaign)	Lower Confidence Bound(campaign)	Upper Confidence Bound(campaign)
2015	5368			
2016	8308			
2017	15180			
2018	12332	12332	12332.00	12332.00
2019		16218.60534	11560.75	20876.46

Statistic	Value
Alpha	0.10
Beta	0.00
Gamma	0.00
MASE	0.37
SMAPE	0.12
MAE	1,561.50
RMSE	2,376.50

Analysis: The above dashboard explains, estimated count of campaign conducted by banks in the year 2015 was 5368. But, in 2016, the count was around 8308 which is higher than the previous year. As the year increases, there is a huge difference in numbers. The values of conducting campaigns increased and is almost double from last year. The very next year there is a slight decrease in the numbers as there was not much response recorded from the customers which lead to the declination of campaigns conducted. But the trend shows that there will be a greater number of campaigns in future as bank still predicts from historical, values that there will be more customer responses and they might subscribe for term deposits.

Dashboard 17:



Predictive Forecasting:

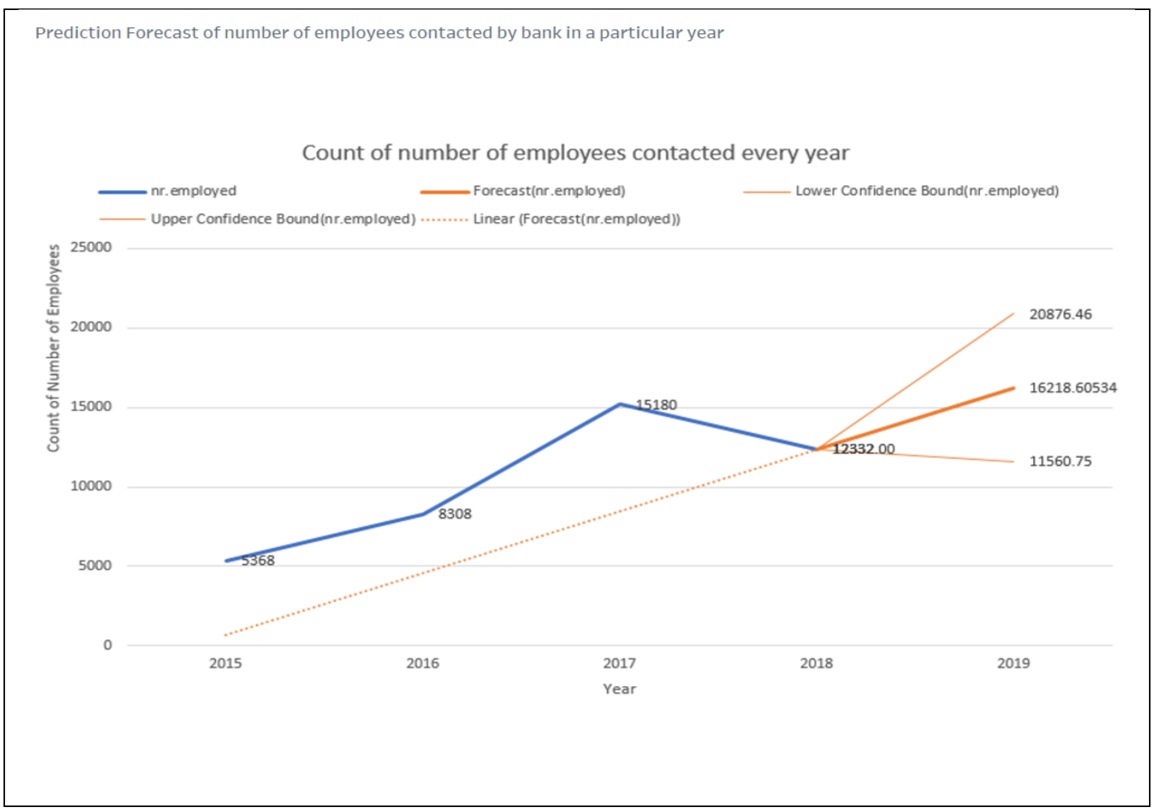
Time series:	Year
Measures:	Campaign
Forecast forward:	2 year (2018 – 2020)
Forecast based on:	2015 – 2018
Confidence Interval	95%
Seasonality	As the data is yearly, we cannot support seasonality option over here.

year	age	Forecast(age)	Lower Confidence Bound(age)	Upper Confidence Bound(age)
2015	40.64083			
2016	39.62506			
2017	40.3135			
2018	39.6681	39.66809925	39.67	39.67
2019		39.48115009	38.65	40.31
2020		39.19283776	38.35	40.04

Statistic	Value
Alpha	0.10
Beta	0.10
Gamma	0.00
MASE	0.38
SMAPE	0.01
MAE	0.30
RMSE	0.42

Analysis: From the above graph, it can be said that from the year 2015-17, the average age group of customers contacted was fluctuating and after 2018 age group of young people were contacted more. This is because bank predicted from historical value that young age group of people were more interested in subscription of term deposits.

Dashboard 18:



Prediction Forecast:

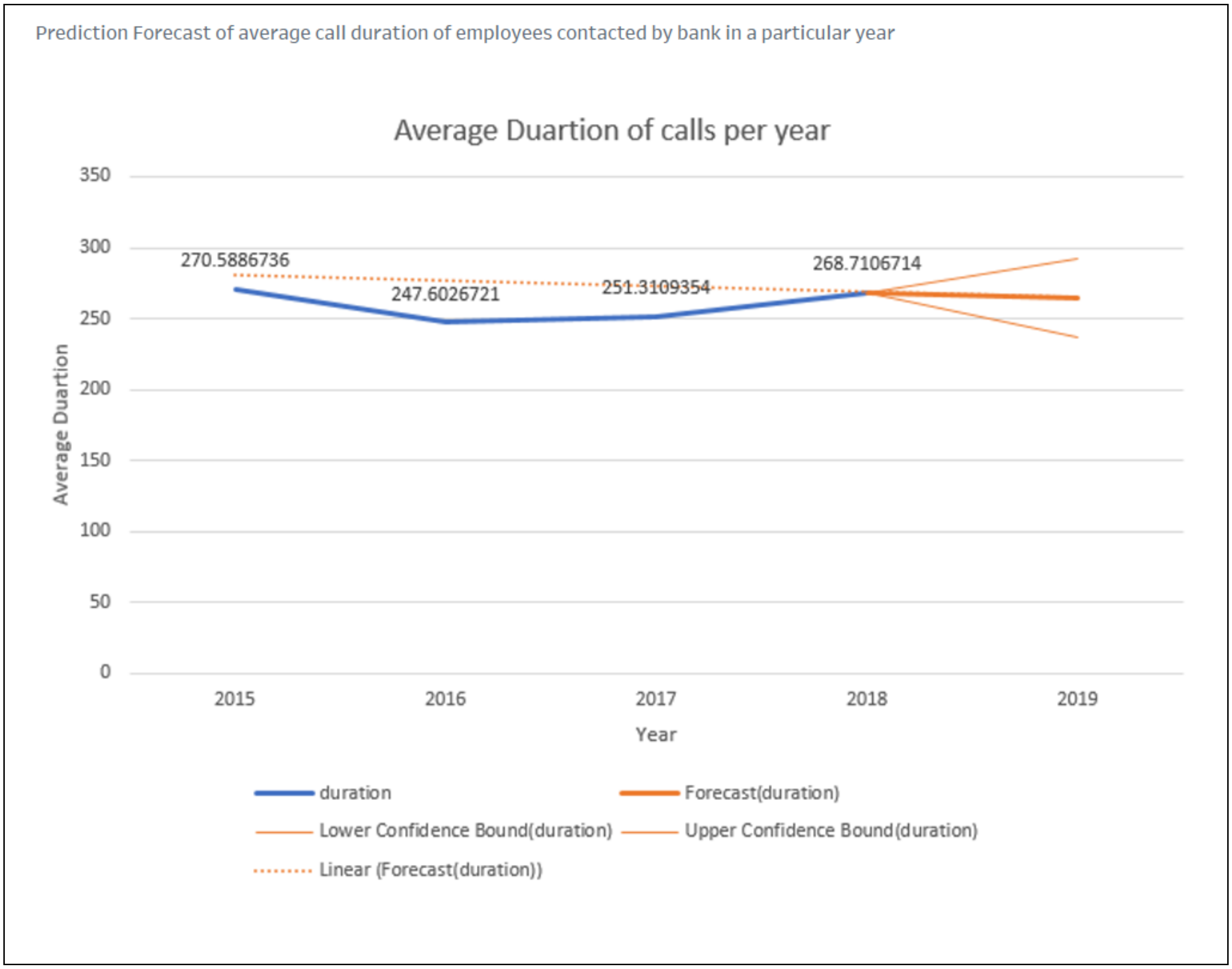
Time series:	Year
Measures:	Campaign
Forecast forward:	1 year (2018 – 2019)
Forecast based on:	2015 – 2018
Confidence Interval	95%
Seasonality	As the data is yearly, we cannot support seasonality option over here.

year	nr.employed	Forecast(nr.employed)	Lower Confidence Bound(nr.employed)	Upper Confidence Bound(nr.employed)
2015	5368			
2016	8308			
2017	15180			
2018	12332	12332	12332.00	12332.00
2019		16218.60534	11560.75	20876.46

Statistic	Value
Alpha	0.10
Beta	0.00
Gamma	0.00
MASE	0.37
SMAPE	0.12
MAE	1,561.50
RMSE	2,376.50

Analysis: In 2015, employed contacted customers were 5k and in next two years it was almost triple (15k). According to historical values it can be predicted, bank have increased their targets of contacting more educated customers to sell their subscription.

Dashboard 19:



Predictive Forecasting:

Time series:	Year
Measures:	Campaign
Forecast forward:	1 year (2018 – 2019)
Forecast based on:	2015 – 2018
Confidence Interval	95%
Seasonality	As the data is yearly, we cannot support seasonality option over here.

year	duration	Forecast(duration)	Lower Confidence Bound(duration)	Upper Confidence Bound(duration)
2015	270.58867			
2016	247.61876			
2017	251.3019			
2018	268.70501	268.7050097	268.71	268.71
2019		264.5051472	236.89	292.12

Statistic	Value
Alpha	0.25
Beta	0.00
Gamma	0.00
MASE	0.75
SMAPE	0.04
MAE	11.00
RMSE	14.09

Analysis: The dashboard defines that, there is very slight difference in the average duration of calls per year. Reason for this could be bank employees explain their policies in same time limit which makes duration to remain constant.

Conclusion:

From the above four dashboards, it can be predicted that the banks are very particular about the increasing their telemarketing campaigns to their customers. The conducts frequent campaigns which helps the customer to think about their promotion on term deposits. Also, they target the customers with following age group and with employment. They call their customers and tries to limit their calling times explaining all the subscription offers on term deposits. Now, it’s up to the clients whether they want to subscribe for term deposits or not.