**Dataset insights :-**

* The data set initially has 41189 observations including the missing values
* The missing values are assumed to be unknown for variables

|  |  |
| --- | --- |
| job | unknown |
| education | unknown |
| marital | unknown |
| loan | unknown |
| housing | unknown |

**Data Preparation:-**

* Missing values Treatment
* I removed these variables to avoid noise in my final result
* After removing the missing values we are left with a total of 38245 observations
* Outliers Treatment
* proc univariate to find out data statistics
* outliers are treatred for the variables are treated individually
* duration to 99th percentile
* Data Conversion :-
* converting all the categorical variable to numerical equivalents
* values are given like 0,1,2,3 depending on their business significance

**Correlation:-**

* made a correlation matrix and found out the variables with high covariance(positive or negative) index coefficient.
* Prediction looking at the correlation matrix

|  |  |
| --- | --- |
| Economic factors are strongly correlated among each other  Previous is negatively correlated to the economic factors  Contact is slightly related P to the economic variables   * Response is correlated to strong positively duration and negatively correlated to “pdays” “previous” “poutcome” | |
|  |
|  | |
| * age is related to marital * Independent Variables:-  |  |  |  |  | | --- | --- | --- | --- | | Month | Day of week | Campaign | Job | | * Education | * Housing | * Loan |  | | |
|  | |

|  |
| --- |
| **Factor Analysis:-**  We observe multi co-linearity among different variables so we try to reduce the no of variables by discarding the not important ones which will otherwise create noise in our final result |
| |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | **Factor1** | **Factor2** | **Factor3** | **Factor4** | **Factor5** | **Factor6** | **Factor7** | **Factor8** | **Factor9** | **Factor10** | | **emp\_var\_rate** | 0.965 | -0.142 | 0.049 | 0.027 | -0.065 | -0.044 | -0.015 | -0.016 | 0.003 | -0.015 | | **euribor3m** | 0.942 | -0.201 | 0.057 | 0.061 | -0.081 | -0.002 | 0.029 | 0.024 | 0.003 | -0.015 | | **nr\_employed** | 0.866 | -0.295 | -0.027 | 0.020 | -0.116 | -0.177 | -0.021 | 0.003 | 0.006 | -0.006 | | **cons\_price\_idx** | 0.820 | 0.111 | 0.079 | -0.042 | 0.043 | 0.252 | -0.124 | -0.115 | 0.001 | -0.040 | | **previous** | -0.310 | 0.882 | -0.220 | 0.010 | 0.053 | 0.015 | 0.031 | 0.024 | 0.015 | -0.012 | | **pdays** | 0.208 | -0.725 | -0.584 | 0.008 | -0.099 | 0.018 | -0.018 | 0.001 | -0.001 | -0.006 | | **poutcome** | 0.147 | -0.074 | 0.940 | -0.021 | 0.063 | -0.028 | -0.012 | -0.033 | -0.016 | 0.020 | | **age** | -0.033 | 0.042 | 0.042 | 0.826 | 0.014 | -0.034 | -0.110 | -0.018 | -0.023 | 0.024 | | **marital** | -0.046 | 0.039 | 0.049 | -0.768 | 0.016 | -0.002 | 0.078 | 0.009 | -0.051 | 0.053 | | **duration** | 0.041 | -0.063 | -0.067 | -0.014 | 0.893 | -0.009 | -0.042 | 0.055 | 0.010 | -0.007 | | **response** | -0.241 | 0.207 | 0.202 | 0.011 | 0.742 | -0.033 | 0.075 | 0.016 | 0.001 | -0.015 | | **month** | -0.197 | 0.025 | -0.066 | -0.045 | -0.023 | 0.814 | -0.080 | 0.072 | 0.016 | -0.005 | | **contact** | 0.483 | -0.034 | 0.082 | 0.030 | -0.023 | 0.708 | -0.046 | -0.076 | -0.012 | -0.053 | | **education** | -0.032 | 0.008 | 0.016 | -0.164 | -0.013 | -0.145 | 0.728 | -0.025 | 0.056 | -0.021 | | **job** | 0.062 | 0.104 | -0.119 | 0.096 | 0.041 | 0.006 | 0.537 | -0.086 | -0.343 | 0.339 | | **cons\_conf\_idx** | 0.227 | 0.004 | 0.334 | 0.312 | -0.006 | 0.247 | 0.405 | 0.197 | -0.025 | -0.034 | | **default** | 0.242 | 0.029 | -0.015 | 0.212 | -0.023 | -0.003 | -0.509 | -0.039 | -0.184 | 0.208 | | **day\_of\_week** | 0.107 | 0.119 | -0.009 | -0.052 | -0.050 | -0.085 | -0.050 | 0.841 | 0.023 | -0.053 | | **campaign** | 0.206 | 0.106 | 0.005 | -0.033 | -0.131 | -0.122 | -0.043 | -0.521 | 0.055 | -0.095 | | **loan** | 0.027 | 0.025 | -0.032 | 0.045 | 0.015 | 0.008 | 0.066 | -0.041 | 0.913 | 0.113 | | **housing** | -0.070 | -0.019 | 0.030 | -0.053 | -0.023 | -0.037 | -0.049 | 0.046 | 0.114 | 0.900 | |
|  |
|  |
|  |
|  |
|  |
|  |

|  |
| --- |
| **Analytics Initiatives** |
|  |
| • **We can find key drivers for revenue or profit by store.** |
|  |
| **Y (DV)= Response** |
|  |
| **X’s (IV’s) = emp\_var\_rate previoous pdays poutcome age marital duration month job day\_of\_week campaign loan housing** |
|  |
| **Technique = Linear Regression** |
| **Since our dependent variable Respone is ina yes or a no of customer being able to buy the product the best suited technique according ton me would be to go for logistic regression** |