Results and README

Part A

1a) We handled the missing values. In our data, we found no missing or null values

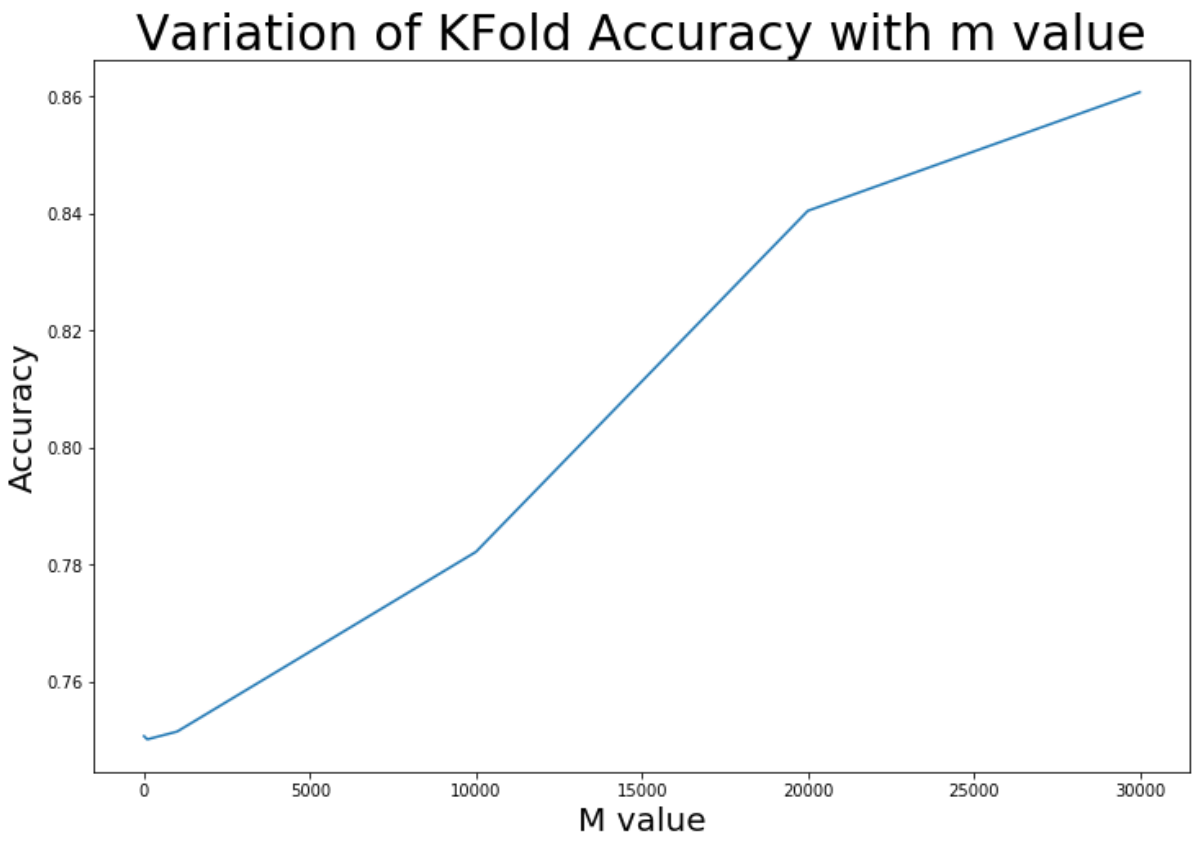
1b) Features encoding done to variables that were considered as Discrete. For the list see report.

1c)

5Fold CV Accuracy (with m = 2): 75.06 %

Test Accuracy (with m = 2): 75.18 %

- We noted that changing the m value in m-estimator method, gave us a increased accuracy
- So, we took m as a hyperparameter and found out the m-value that gave us maximum 5FoldCV accuracy
- Graph of m vs Accuracy attached below:



- We note that the highest 5Fold CV accuracy is for m = 30000. Which gives 85.88 % validation accuracy

Measured Test Accuracy for m = 30000 = 85.55%

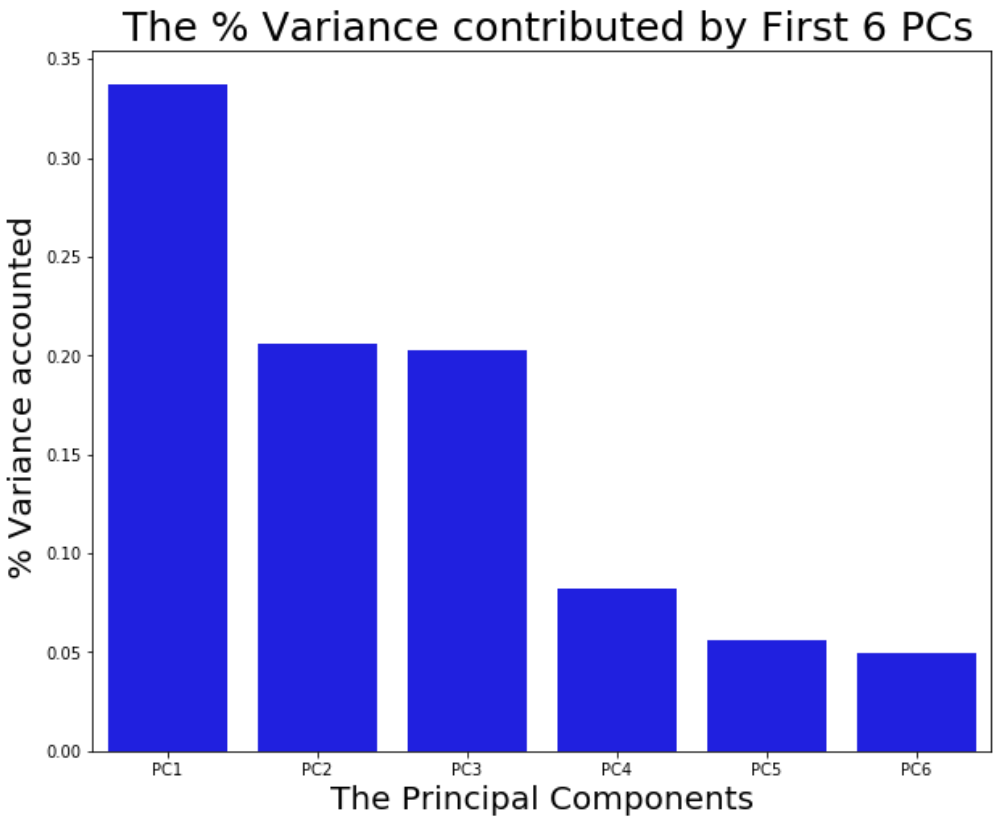
M value	Test Accuracy(%)
2	75.18
100	75.13
1000	75.24
10000	77.41
20000	81.97
30000	85.55

So, we report the Maximum Test Accuracy for Part A as 85.55%

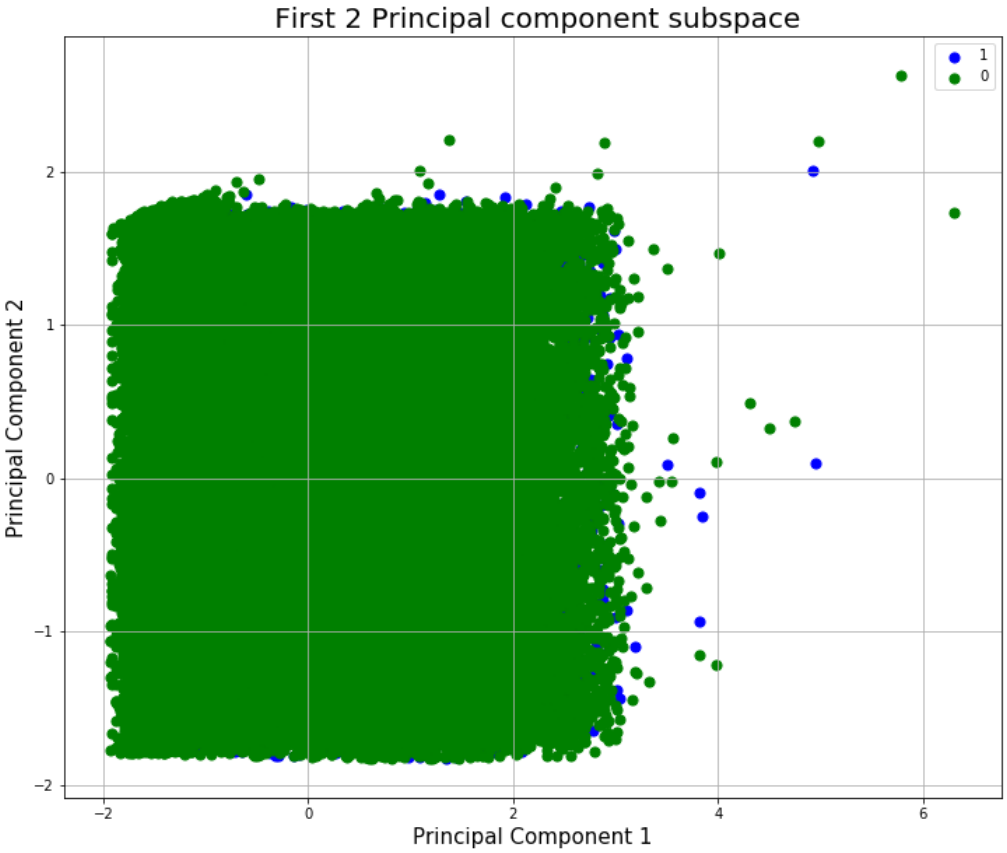
Part B

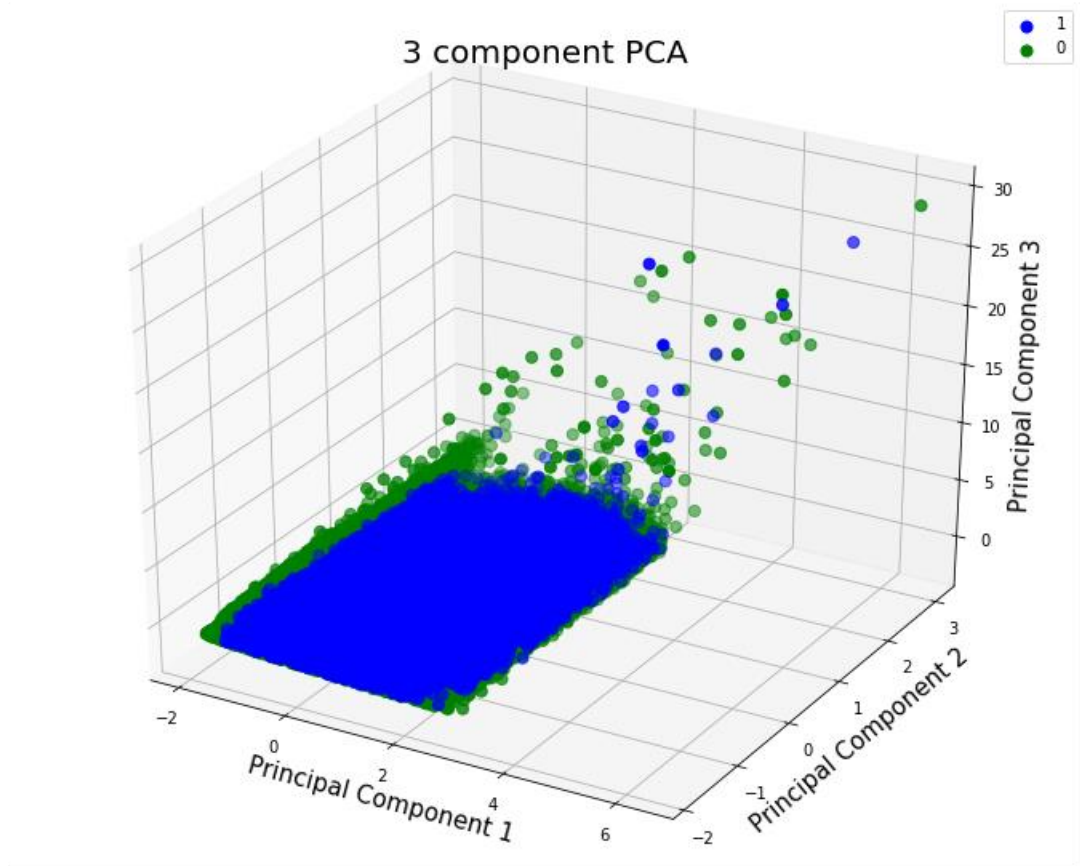
Principal Component Analysis:

- We observed that the First 6 PCs retained about 94% of variance
- So, all results reported are performed using the features as the first 6 PCs of data



2a) PCA Visualisations – Data is Normalised





2b)

5Fold CV Accuracy with PCA features: 85.18 %

Test Accuracy with PCA features: 85.35%

Part 3

- We performed data deletion if the datapoint had more than 30% outlier features.
- Sequential Backward Selection method yielded 7 features:

```
['Gender',
 'Age',
 'Driving_License',
 'Region_Code',
 'Previously_Insured',
 'Policy_Sales_Channel',
 'Vintage']
```

- We note that this method of Backward Feature Selection performed after deleting outliers yielded the best Validation and Test Accuracies

5Fold CV Accuracy with selected features: 87.71 %
Test Accuracy with selected features: 87.86%

Results Conclusion:

Method	Test Accuracy
No feature engineering	75.18 % (m = 1), 85.55% (m = 30000)
PCA features	85.35%
Selected features (Backward)	87.86%

README:

The code execution is self-explanatory, you would just have to follow the comments in the code, All packages mentioned at the start must be installed.

Google Collab Link:
https://colab.research.google.com/drive/1UVgC0jqF1GEb1c5sOmh1JJUv_Lcsuo64?usp=sharing