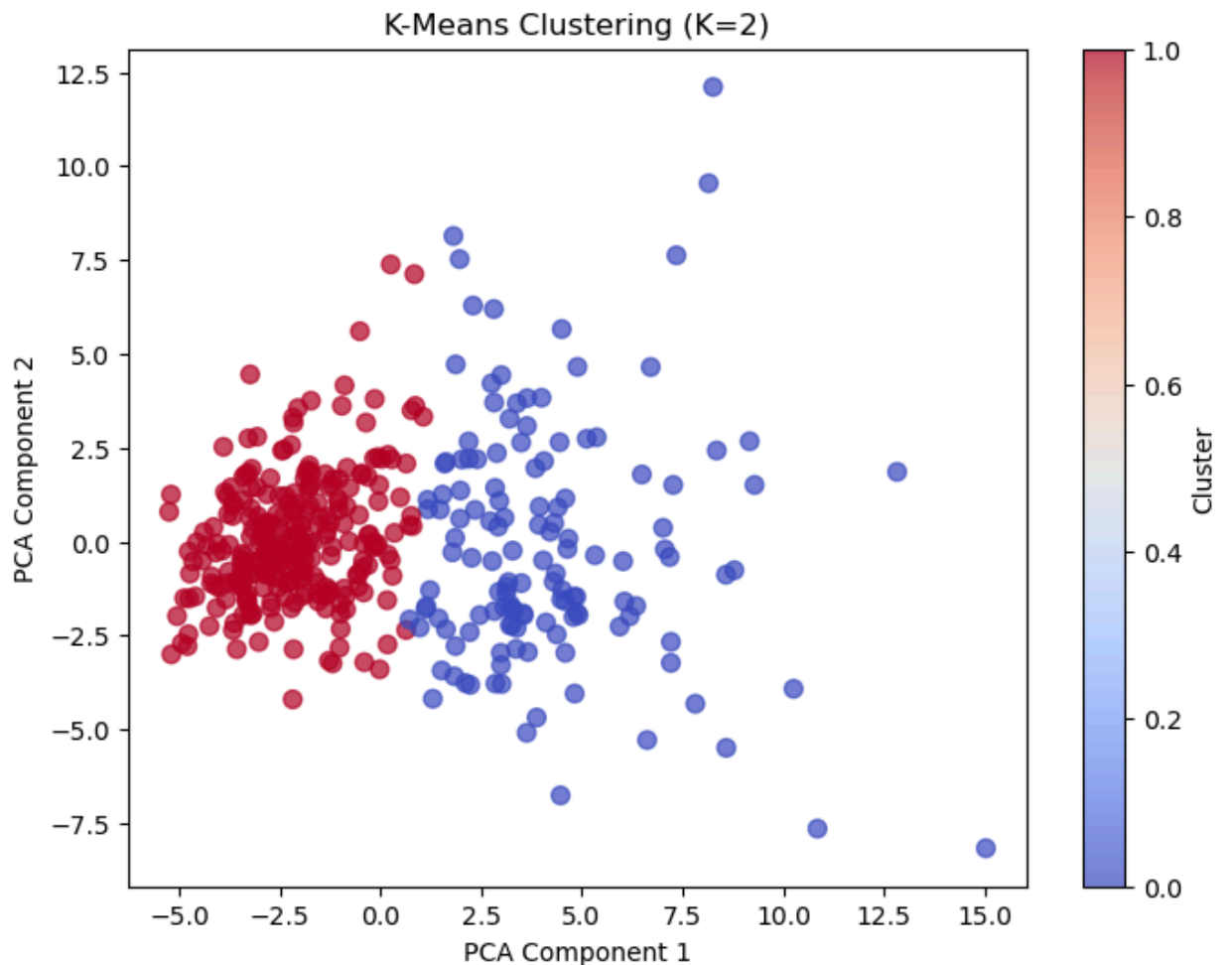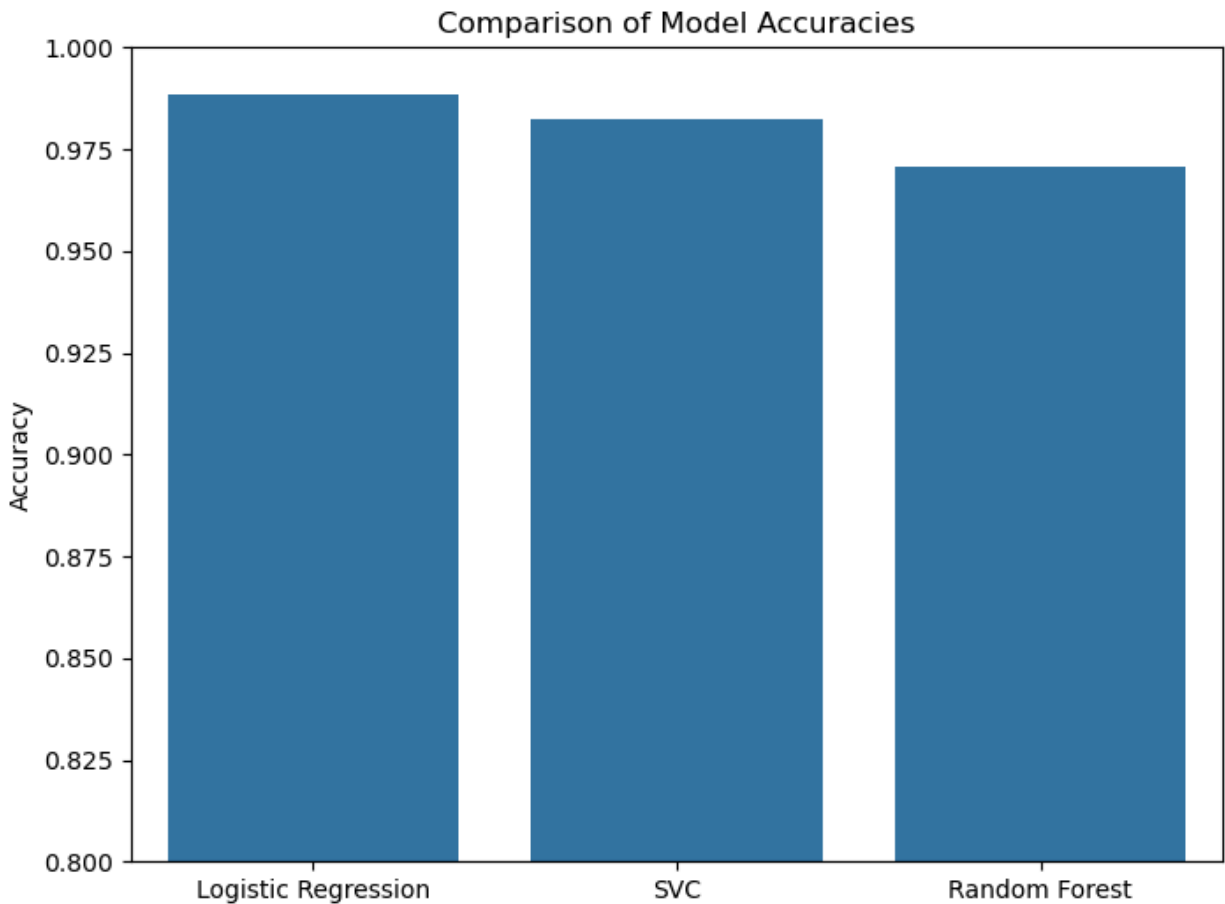**Project Report -2**
Adult Census Data

1. **Include the clustering graph and write your understanding**

- X -axis shows the first principle component where it holds the largest amount of variance in the data .
- Y- axis shows the second principle component where it holds the second largest variance .

- From the cluster we can say that the red cluster has the tight distribution that is lower PCA component value.
- The blue cluster is more dispersed shows the higher variance .

2. **Discuss the accuracies of all the classification models implemented with the help of the comparison graph.**

Comparison of Model Accuracies

- Accuracy of Logistic Regression **98%**
- Accuracy of SVC **98%**
- Accuracy of Random Forest **97%**
- Logistic regression give a linear relationship between features and target variable .
- It show the dataset performs well in logistic regression .
- SVC is same as logistic regression what it make sense is it captures the non linear relationship in the data . So it makes the model perform well in linear as well as non linear .

- Random forest perform slightly lower than other two models. Because it performs well in larger data . However 97 % accuracy is good and it also predict correctly . In some causes it may fail .

3. Which model do you think performed well and why?

- Among the three models logistic regression performs well in terms of accuracy . It correctly relate the relationship between features and target .

- In addition to this SVC is also gives better result when it is non liner relationship . It handles non linear patter in the data .