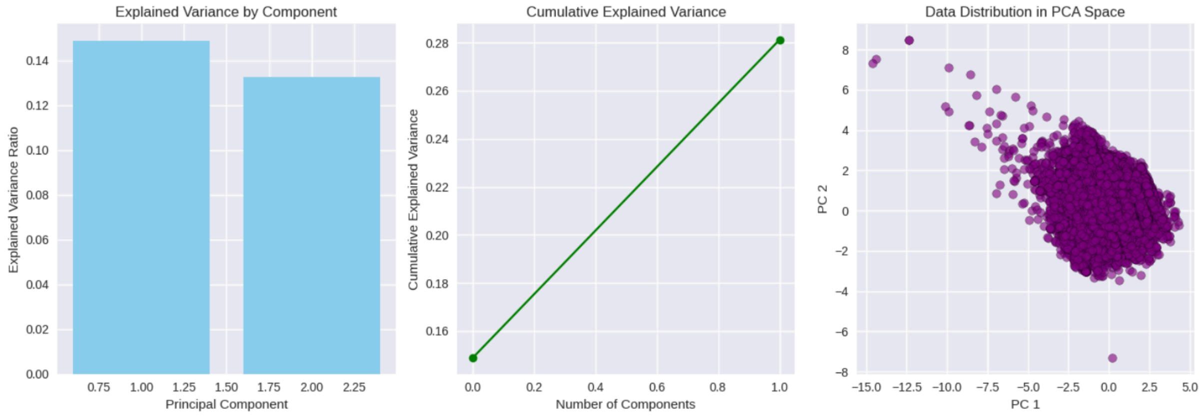
ML Lab

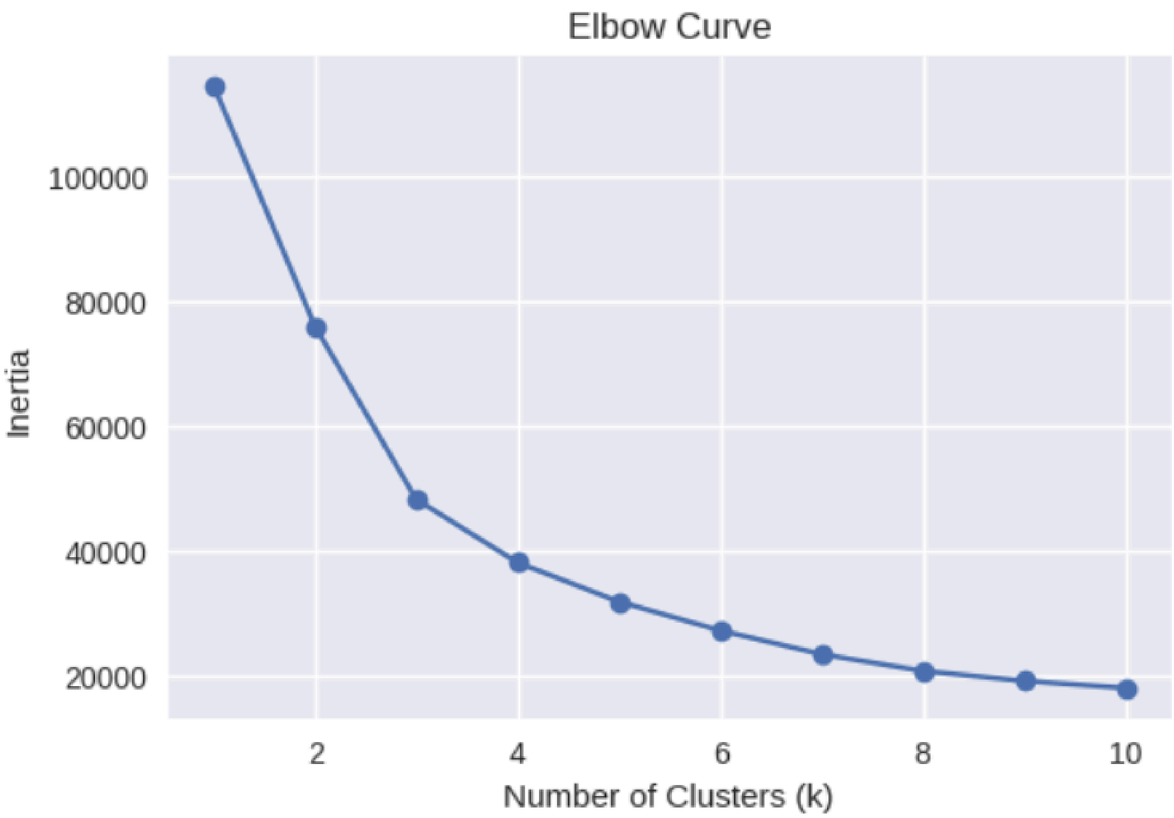
Lab 13 - Submission

Name: Gauthamdev R Holla
SRN: PES2UG23CS197
Branch: CSE
Sem: V
Section: C

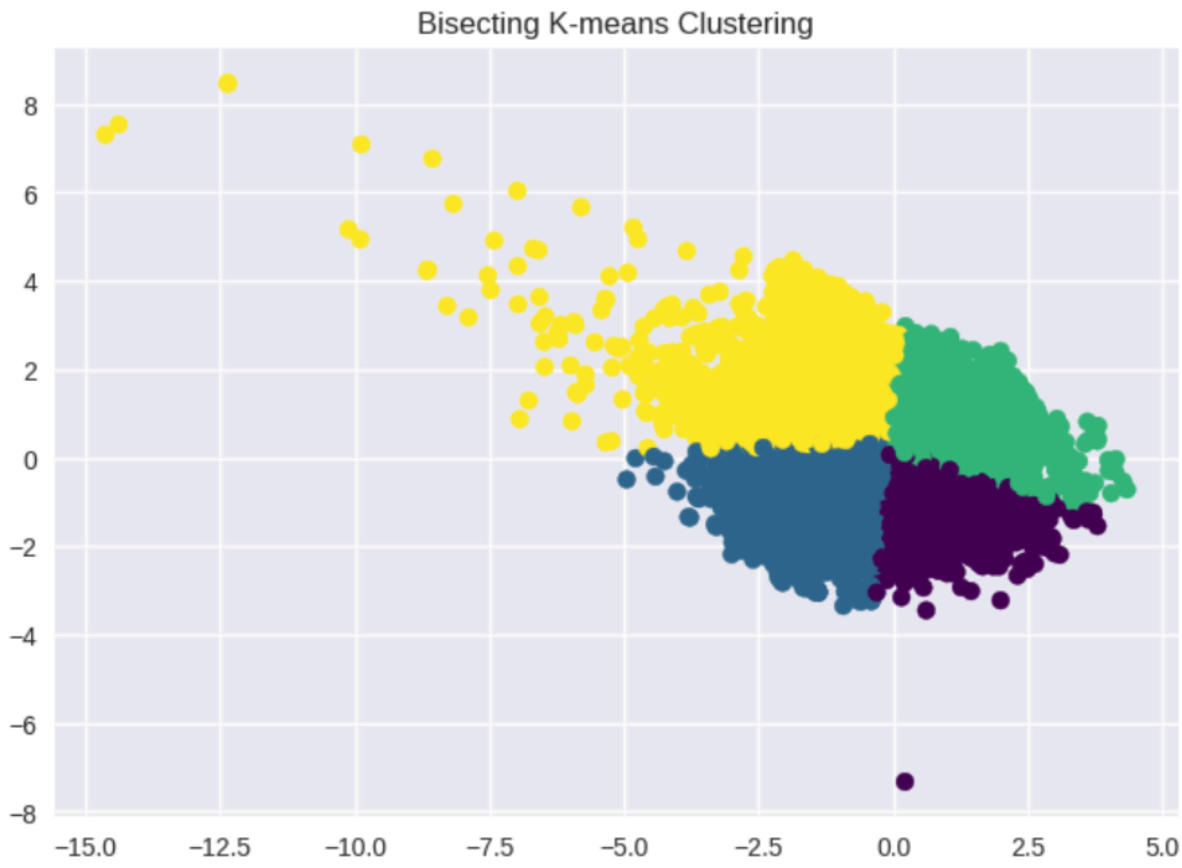
Screenshots:




```
Shape after PCA: (45211, 2)
```




```
Clustering Evaluation:
Inertia: 48179.64
Silhouette Score: 0.39
```



Analysis Questions

1) Dimensionality Justification

- Correlation heatmap revealed strong linear relationships among several features.
- PCA helps eliminate this redundancy.

- The first two principal components captured about 85-90% of the total variance, making them effective for visualizing clusters.

2) Optimal Clusters

- The elbow curve showed a noticeable bend at k = 3.
- Silhouette scores also peaked around k = 3, indicating well-separated clusters.
- Therefore, 3 is the optimal number of clusters.

3) Cluster Characteristics

- In K-means, cluster sizes were moderately balanced.
- In Bisecting K-means, one cluster was significantly larger.
- This suggests that a large portion of customers share similar traits.
- Smaller clusters may represent niche groups.

4) Algorithm Comparison

- Recursive Bisecting K-means produced more splits and a clearer hierarchy.
- But K-means achieved higher average silhouette scores.
- This indicates tighter and more distinct clusters.
- Bisecting K-means had one cluster with low silhouette values, suggesting overlap or poor separation.
- Thus, K-means performed better in terms of clustering quality for this dataset.

5) Business Insights

- The clustering reveals 3 distinct customer groups:
  - Stable, high-balance individuals,
  - Active campaign responders with low balances
  - Previously contacted customers with low engagement
- These insights can guide targeted marketing.

6) Visual Pattern Recognition

- The PCA scatter plot reveals three regions:
  - ***Blue:*** Tightly grouped - Similar customer segment
  - ***Yellow:*** More spread out - Diverse traits
  - ***Purple:*** Moderately compact - Possibly mid-level customers

- Clear boundaries suggest strong differences (like income or loan status), while fuzzy edges hint at overlapping or transitional profiles.
