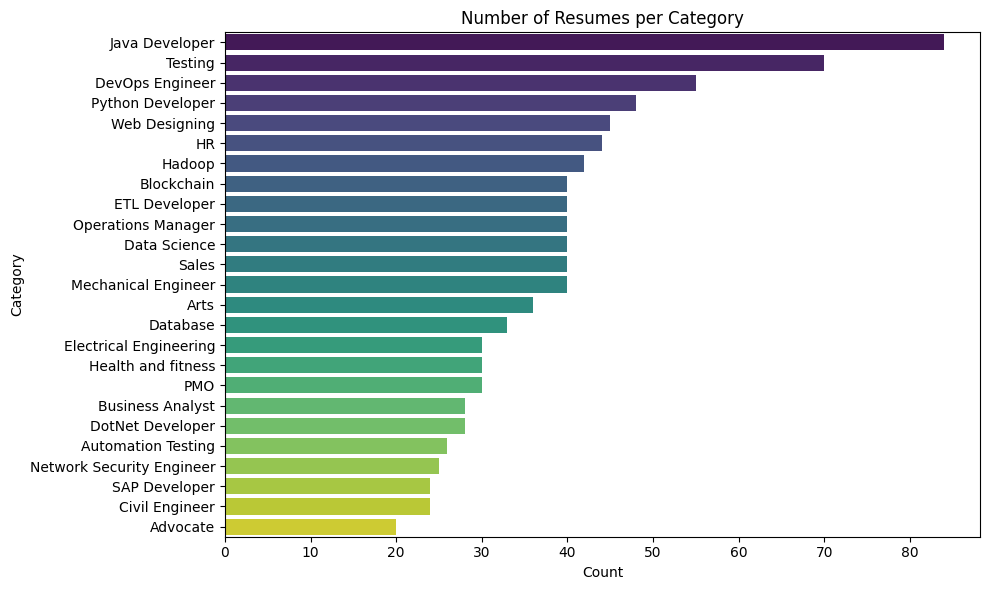
Name – Aryan Kakran

**REPORT**

**Summary and Visualisation**

****

**1: Count of Resumes per Category**

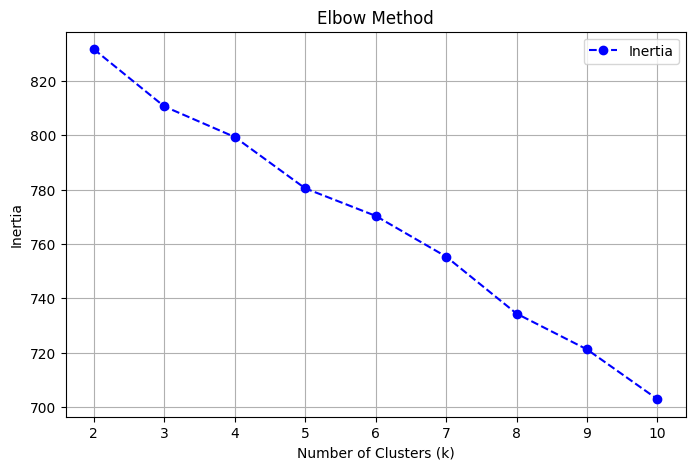
* Figure Description: Bar plot displaying the number of resumes for each category.
* Purpose: Understand the distribution of resumes across categories.
* Insight: Highlights which categories have the most or least resumes, helping identify focus areas.

A graph of a number of columns

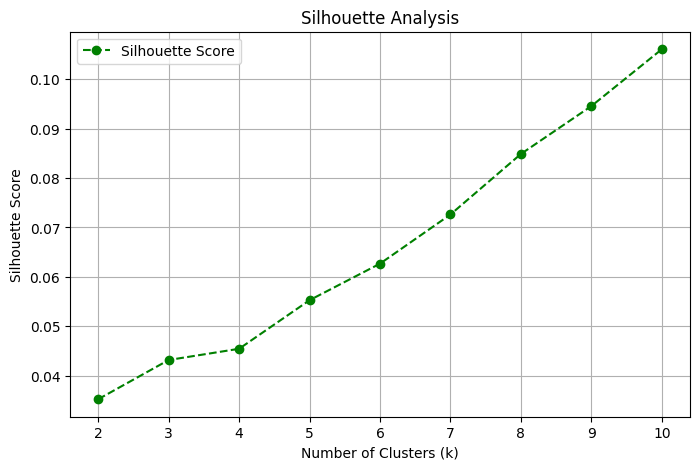
Description automatically generated

**2: Distribution of Word Count in Resumes**

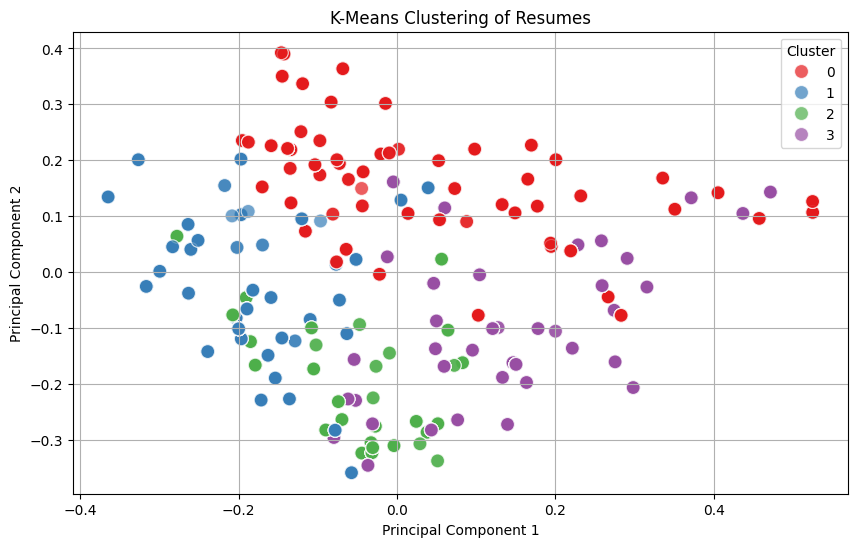
* Figure Description: Histogram showing the distribution of word counts across all resumes.
* Purpose: Analyse the length of resumes to identify patterns in word usage.
* Insight: Provides a general overview of how verbose resumes are and identifies potential outliers.

**3: Elbow Method (Inertia)**

* Figure Description: Line plot showing the inertia for different numbers of clusters.
* Purpose: Identify the optimal number of clusters for K-means clustering.
* Insight: The "elbow" point in the graph suggests the best number of clusters to balance compactness and simplicity.

**4: Silhouette Analysis**

* Figure Description: Line plot showing silhouette scores for different numbers of clusters.
* Purpose: Evaluate the quality of clustering for each cluster count.
* Insight: Higher silhouette scores indicate better-defined clusters.



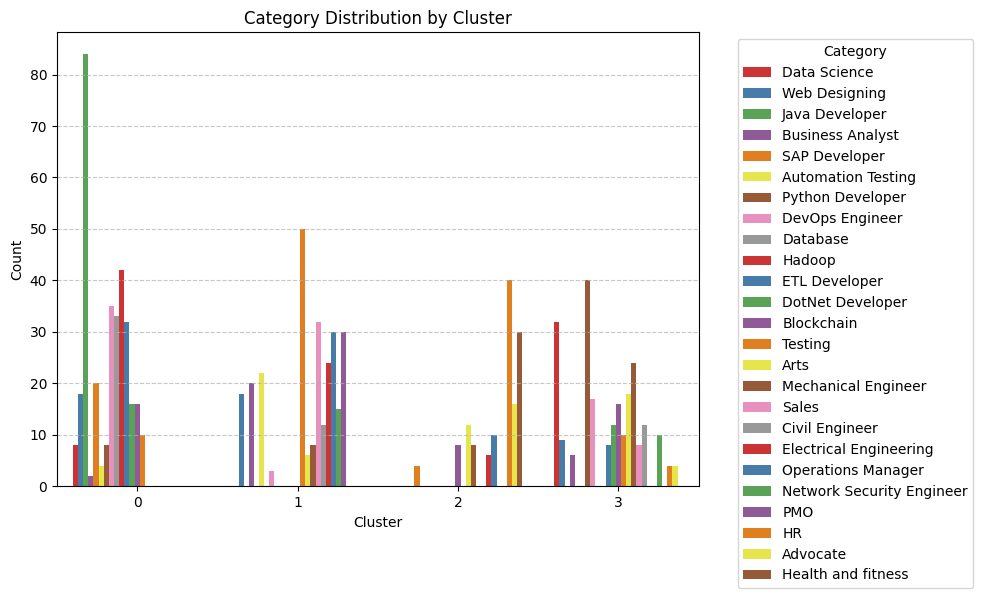
**5: K-Means Clustering Visualization**

* Figure Description: Scatter plot of resumes reduced to 2D using PCA, coloured by cluster.
* Purpose: Visualize how resumes are grouped into clusters.
* Insight: Showcases the separability and overlap between clusters.

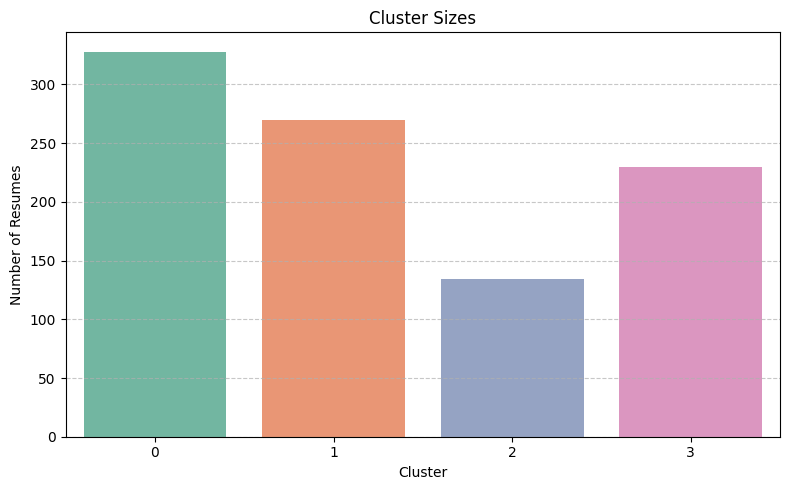
**A close up of words

Description automatically generated6: WordCloud for Each Cluster**

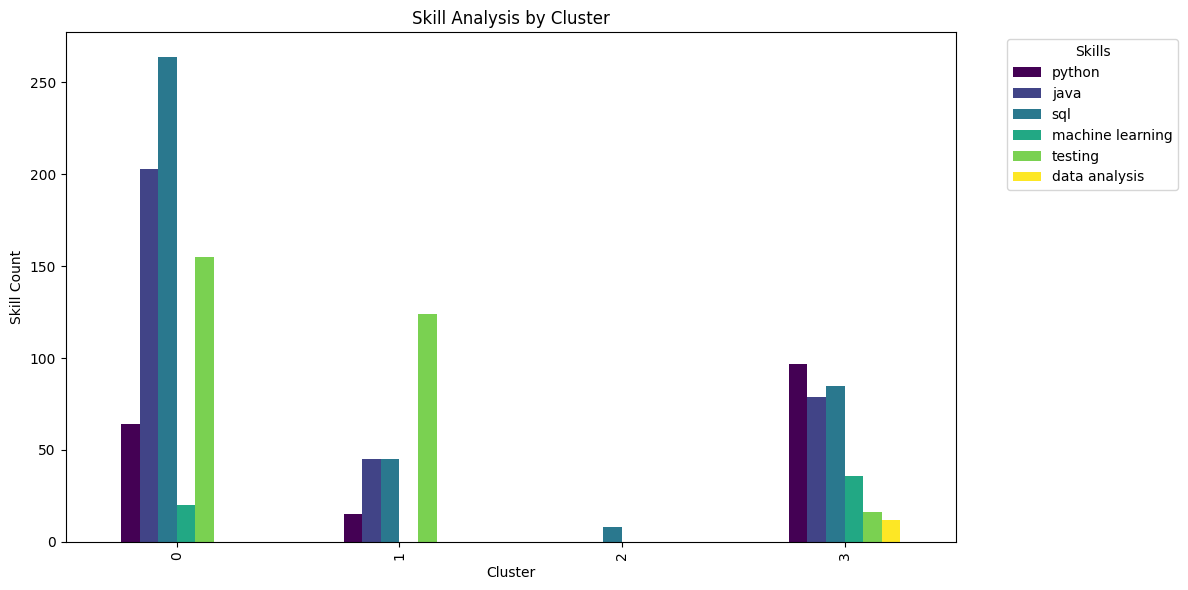
* Figure Description: WordCloud representing frequent terms in each cluster.
* Purpose: Highlight important keywords and skills for each cluster.
* Insight: Reveals focus areas or expertise represented by resumes in each cluster.

**7: Category Distribution by Cluster**

* Figure Description: Stacked bar plot showing the distribution of categories in each cluster.
* Purpose: Analyze how well categories align with clusters.
* Insight: Helps assess whether clustering reflects the inherent structure of the categories.

**8: Cluster Sizes**

* Figure Description: Bar plot showing the number of resumes in each cluster.
* Purpose: Evaluate the balance of clusters.
* Insight: Highlights any imbalances, such as clusters with too few or too many resumes.

**9: Skill Analysis by Cluster**

* Figure Description: Bar plot showing the count of key skills in each cluster.
* Purpose: Identify the skill distribution across clusters.
* Insight: Provides a detailed view of skill prevalence in different clusters, helping align clusters with desired competencies.