1. Project Overview

The objective of this project is to perform customer segmentation using clustering techniques and to build prediction models for customer churn. The project helps businesses understand different customer behaviors and apply segment-specific strategies to improve customer retention and revenue.

---

2. Dataset Description

The dataset contains customer information including tenure, monthly charges, total charges, contract type, payment method, billing preference, and churn status. Each row represents a unique customer.

---

3. Tools and Technologies Used

- Python
- Jupyter Notebook
- Pandas and NumPy
- Scikit-learn
- Matplotlib and Seaborn

---

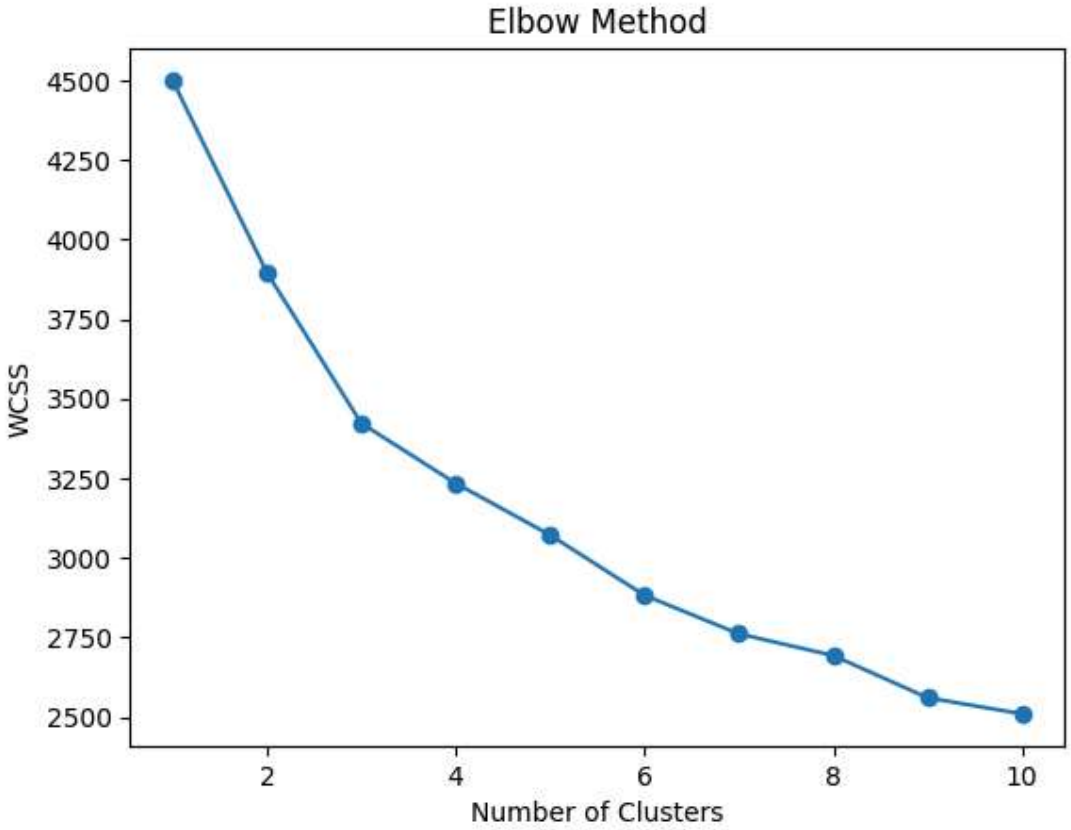
4. Methodology

4.1 Data Preprocessing

- Removed identifier column (CustomerID)
- Handled categorical variables using one-hot encoding
- Scaled numerical features using StandardScaler

4.2 Clustering Techniques

- K-Means clustering with Elbow Method



-
- K-Means clustering was applied to segment customers based on their behavior.
- To determine the optimal number of clusters, the Elbow Method was used.
-
- The Elbow Method plots the number of clusters against the Within-Cluster Sum of Squares (WCSS).
- The point where the curve shows a clear bend (elbow) indicates the optimal number of clusters.
-
- Based on the Elbow Method graph (shown above), the optimal number of clusters was selected as K = 3.
-
- In addition to K-Means, Hierarchical Clustering and DBSCAN were applied to compare clustering performance and validate the segmentation results.

4.3 Segment Profiling

Clusters were analyzed based on average values of tenure and charges. Each cluster was assigned a meaningful business name.

4.4 Prediction Models

Random Forest classifiers were trained separately for each customer segment to predict churn.

4.5 Model Evaluation

Models were evaluated using:

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC

---

5. Business Implications

Customer segmentation allows businesses to:

- Design targeted marketing strategies
- Reduce churn through personalized offers
- Improve customer satisfaction and retention
- Increase overall revenue

---

6. Results Summary

The models achieved high accuracy across different segments. Premium Spenders showed the highest retention, while Budget Conscious customers required focused retention strategies.

---

7. Limitations

- Dataset size is limited
- Customer behavior assumed to be static
- External factors were not included

---

8. Future Enhancements

- Use real-time data
- Apply advanced ML models
- Include demographic features
- Deploy as a web application

---

9. Conclusion

This project demonstrates how clustering and machine learning models can be effectively used to derive business insights and improve decision-making in customer management.