# Customer Churn Prediction & Segmentation with Python & Power BI

This end-to-end project identifies potential churners in a bank’s customer base using machine learning, and visualizes actionable insights through an interactive Power BI dashboard.

## Project Overview

Many businesses struggle to retain customers. This project focuses on predicting customer churn and segmenting users for targeted retention strategies using both Python and Power BI.

* Cleaned and preprocessed customer data
* Performed exploratory data analysis (EDA)
* Built machine learning models to predict churn
* Handled class imbalance using **SMOTE**
* Applied **KMeans clustering** for customer segmentation
* Reduced dimensions using **PCA** for cluster visualization
* Built an interactive dashboard in **Power BI**

## Files Included

* customer\_churn\_prediction\_with\_segmentation.ipynb: Jupyter Notebook with full pipeline (EDA ➝ ML ➝ Clustering)
* Bank Customer Churn Prediction.csv: Dataset used for training and analysis
* customer churn dashboard.pbix: Final interactive Power BI dashboard

## Technologies Used

| Tool | Purpose |
| --- | --- |
| Python | Programming Language |
| Pandas, NumPy | Data cleaning & manipulation |
| Seaborn, Matplotlib | Data visualization |
| Scikit-learn | ML models: Logistic, RF, KMeans |
| SMOTE | Handle class imbalance |
| PCA | Dimensionality Reduction |
| Power BI | Data Visualization Dashboard |

## Exploratory Data Analysis (EDA)

* Churn distribution
* Age vs Churn rate
* Geography and gender insights
* Tenure vs Balance correlation

## Machine Learning Models

* **Logistic Regression**
* **Random Forest Classifier**

**Evaluation Metrics:**

* Accuracy
* ROC-AUC
* Classification Report

## Customer Segmentation

Used **KMeans clustering** to group customers into meaningful segments:

* High-risk churners
* Loyal customers
* Low engagement customers

Applied **PCA** for 2D visualization of clusters.

## Power BI Dashboard Highlights

* **KPIs:** Total Customers, Churn Rate, Retention %
* **Filters:** Gender, Geography, Tenure, Age Range
* **Visuals:** Clustered column charts, line charts, churn by age & tenure
* **Goal:** Help stakeholders monitor churn risk and segment behavior

## How to Use

1. Open the Jupyter notebook and run each cell to reproduce the analysis
2. Load the .pbix file in Power BI Desktop to explore the dashboard
3. Replace the dataset with your own CSV for reuse

## Key Results

* Achieved **87% accuracy** with Random Forest
* Applied **SMOTE** to improve model performance
* Built clear visual stories using **Power BI**
* Provided segmentation-based recommendations for marketing

## Author

**Veera Sai Pavan Chavvakula**

* [veerasaipavan6673@gmail.com](mailto:veerasaipavan6673@gmail.com)
* [LinkedIn](https://www.linkedin.com/in/veera-sai-pavan-chavvakula-6260a72bb) : https://www.linkedin.com/in/veera-sai-pavan-chavvakula-6260a72bb
* [GitHub](https://github.com/veerasaipavan): <https://github.com/veerasaipavan>