Summary of Telecom Customer Churn Analysis & Logistic Regression Model

1. Exploratory Data Analysis (EDA) - Graph Insights

During EDA, various graphs were used to analyze customer churn patterns. Here are the key takeaways:

a. Churn Distribution (Pie Chart & Countplot)

- The **percentage of churned customers** was visualized.
- Churn is significantly lower compared to non-churn, indicating **class imbalance**.
- A countplot showed more customers did **not churn** (0) compared to those who **did churn** (1).

b. Gender vs. Churn

- The churn rate was **similar for both genders**, suggesting gender is **not a strong predictor**.

c. Senior Citizens & Churn

- **Senior citizens had a higher churn rate** than younger customers.
- This suggests age-based customer retention strategies might be useful.

d. Monthly Charges vs. Churn

- Higher **monthly charges** were associated with **higher churn**.
- Customers with lower monthly charges tended to stay.

e. Tenure vs. Churn

- Customers with **longer tenure** had **lower churn rates**.
- New customers were more likely to churn, highlighting the importance of **onboarding strategies**.

---

2. Logistic Regression Model Performance

After preprocessing and training the logistic regression model, the following results were obtained:

a. Accuracy: 80.1%

- The model correctly classified **80.1% of the cases**.
- This is a good baseline but might be improved.
-
-

- **Confusion Matrix**

```
[922 113]
[167 207]]
```

- **922 True Negatives (TN):** Correctly predicted as "Not Churn."
- **113 False Positives (FP):** Wrongly predicted as "Churn" (non-churn customers).
- **167 False Negatives (FN):** Customers who **actually churned but were predicted as non-churn** (this is problematic).
- **207 True Positives (TP):** Correctly predicted as "Churn."

c. Precision, Recall, and F1-Score

Class	Precision	Recall	F1-Score
Non-Churn (0)	**85%**	**89%**	**87%**
Churn (1)	**65%**	**55%**	**60%**

Observations:

- **High precision for Class 0 (Non-Churn):** 85% of customers predicted as **not churning** were correctly classified.
- **Low recall for Class 1 (Churn):** The model **missed 45% of actual churners**.
- The **churn class is under-predicted**, indicating the model is biased toward non-churn cases.

Final Thoughts

- The model is **good at identifying non-churners** but **struggles to correctly classify churned customers**.
- To **reduce customer churn**, the company should:
    - Improve retention efforts for **new customers and senior citizens**.
    - Offer **discounts on high monthly charges** to reduce churn risk.
    - Implement **predictive customer outreach** based on high-risk churn profiles.