Churn Analysis Report

1. Introduction

Customer churn (the percentage of customers who stop using a service) is one of the biggest challenges for subscription-based businesses. This project focuses on analyzing customer data to identify factors contributing to churn and developing a **churn prediction model** using **Random Forest Classifier**. Additionally, dashboards were created to monitor churn patterns and track key performance indicators (KPIs).

---

2. Dataset

- **Source:** Telecom customer dataset (CSV)
- **Records:** ~7,000+ customer entries
- **Features:** Demographics (gender, age, dependents), account details (tenure, contract type, payment method), services subscribed (phone, internet, streaming, etc.), and churn status.

---

3. Tools & Technologies

- **Languages:** Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn)
- **Database:** SQL (for querying and preprocessing)
- **Visualization:** Power BI / Excel dashboards
- **Version Control:** Git, GitHub

---

4. Process

1. **Data Cleaning & Preprocessing**
    - Handled missing values and duplicates
    - Converted categorical variables to numerical using encoding
    - Standardized features for ML model input
2. **Exploratory Data Analysis (EDA)**
    - Univariate & bivariate analysis
    - Visualization of churn vs non-churn customers

    - Identification of key churn drivers (tenure, contract type, monthly charges, etc.)

3. **Dashboard Development**
    - Created **Summary Dashboard** (customer distribution, churn ratio)
    - Created **Detailed Dashboard** (demographics, contract type, payment method analysis)
4. **Machine Learning Model**
    - Split dataset into train/test sets (80/20)
    - Applied **Random Forest Classifier**
    - Evaluated using **accuracy, precision, recall, and F1-score**
    - Achieved **~80%+ accuracy** in predicting churn

---

5. Key Insights

- Customers with **month-to-month contracts** churned more compared to long-term contracts.
- **High monthly charges** increased churn probability.
- **Electronic check payment method** had the highest churn rate.
- Customers with **short tenure (<12 months)** were most likely to churn.
- Fiber optic internet users showed higher churn than DSL users.

---

6. Conclusion

The analysis shows that **contract type, tenure, and monthly charges** are the most significant churn drivers. The Random Forest model helps predict potential churners with high accuracy, allowing businesses to proactively take retention measures.

Recommendations:

- Promote **long-term contracts** with discounts to reduce churn.
- Provide **personalized offers** for high-charge customers.
- Improve **customer support** for fiber optic users.
- Target **early-tenure customers** with onboarding benefits to increase loyalty.