**Exploratory Data Analysis (EDA) with Pandas in Customer Churn Analysis**

The project aims to analyse, predict, and reduce customer churn in a telecom company using data analysis. It identifies key factors influencing churn, provides business insights, and helps develop customer retention strategies to improve profitability and customer satisfaction**.**

**Goals of the Project**

* Understand customer churn patterns and identify key factors influencing churn.
* Perform univariate and bivariate analysis to explore customer behaviour.
* Use data visualization to uncover insights.
* Apply feature engineering to improve model performance.
* Predict churn using machine learning models.
* Provide actionable recommendations to reduce churn.

**Materials and Methods**

**Materials**

* **Dataset**: Telco-Customer-Churn.csv (Telecom company customer data).
* **Libraries**:
* Pandas → Data handling
* NumPy → Numerical analysis
* Seaborn & Matplotlib → Data visualization
* **Tools**: PyCharm

**Methods**

1. **Data Preprocessing**
   * Handle missing values, duplicates, and incorrect data types.
   * Convert categorical variables into numerical formats.
2. **Exploratory Data Analysis (EDA)**
   * Univariate and bivariate analysis.
   * Identify trends and correlations.
3. **Feature Engineering**
   * Create new meaningful features.
   * Standardize and encode categorical features.
4. **Model Building & Evaluation**
   * Train and test machine learning models (e.g., Logistic Regression, Decision Tree, Random Forest).
   * Evaluate using accuracy, precision, recall, and F1-score.

**General Part**

* **Definition of Churn**: Customers who stop using a service.
* **Business Impact**: High churn means loss of revenue and higher acquisition costs.
* **Importance of Analysis**: Helps predict churn and design retention strategies.

**Project Outcome**

* Successfully analysed customer churn patterns using data visualization and statistical analysis.
* Identified key factors influencing churn, such as contract type, payment method, tenure, and monthly charges.
* Developed predictive models to estimate the likelihood of a customer churning.
* Provided data-driven recommendations to improve customerretention

**Key Insights from the Analysis**

**Churn Rate Analysis**

* Overall churn rate: ~26% (indicating a significant loss of customers).
* A large proportion of churned customers had month-to-month contracts.

**Factors Influencing Churn**

1. Contract Type Matters
   * Customers with month-to-month contracts had a higher churn rate.
   * Yearly contract customers were less likely to churn due to commitment.
2. Payment Method Impact
   * Electronic check users had the highest churn rate, possibly due to perceived inconvenience.
   * Customers using automatic bank transfers had the lowest churn rate.
3. Senior Citizens Churn More
   * Older customers (SeniorCitizen = 1) showed higher churn rates, likely due to service usability concerns.
4. Tenure and Churn Relationship
   * Customers with a short tenure (0–12 months) had high churn rates.
   * Loyal customers (tenure > 24 months) were significantly less likely to churn.
5. Monthly Charges Influence Churn
   * Customers with higher monthly charges were more likely to churn, indicating cost sensitivity.

**Feature Engineering**

* **Convert categorical variables** (e.g., Yes/No → 1/0).
* **Create new features**:
  + **Total Charges per tenure** (Average spending per month).
  + **Customer tenure category** (Short-term, Mid-term, Long-term).
* **Normalize numerical features** to improve model performance.

**Key Questions**

1. **What is the distribution of tenure?**

sns.histplot(df["tenure"], bins=30, kde=True, color="#1f77b4")

plt.title("Distribution of Tenure")

plt.xlabel("Tenure (Months)")

plt.ylabel("Count")

plt.show()

1. **What does the histogram of Monthly Charges indicate?**

sns.histplot(df["MonthlyCharges"], bins=30, kde=True, color="#ff7f0e")

plt.title("Distribution of Monthly Charges")

plt.xlabel("Monthly Charges ($)")

plt.ylabel("Count")

plt.show()

1. **What insight can we gain from the Total Charges distribution?**

sns.histplot(df["TotalCharges"], bins=30, kde=True, color="#2ca02c")

plt.title("Distribution of Total Charges")

plt.xlabel("Total Charges ($)")

plt.ylabel("Count")

plt.show()

1. **What is count plot for Contract types?**

ax = sns.countplot(x="Contract", data=df, palette="coolwarm")

ax.bar\_label(ax.containers[0])

plt.title("Count of Contract Types")

plt.xticks(rotation=45)

plt.show()

1. **What is Count of Internet Service Types**

plt.figure(figsize=(6, 4))

ax = sns.countplot(x="InternetService", data=df, palette="coolwarm")

ax.bar\_label(ax.containers[0])

plt.title("Count of Internet Service Types")

plt.xticks(rotation=45)

plt.show()

1. **What is Count of Payment Methods**

plt.figure(figsize=(7, 4))

ax = sns.countplot(x="PaymentMethod", data=df, palette="coolwarm")

ax.bar\_label(ax.containers[0])

plt.title("Count of Payment Methods")

plt.xticks(rotation=45)

plt.show()

1. **What is Churn Distribution**

ax = sns.countplot(x = 'Churn', data = df,palette="coolwarm")

ax.bar\_label(ax.containers[0])

plt.title("Count of Customers by Churn")

plt.show()

1. **What is churn Percentage**

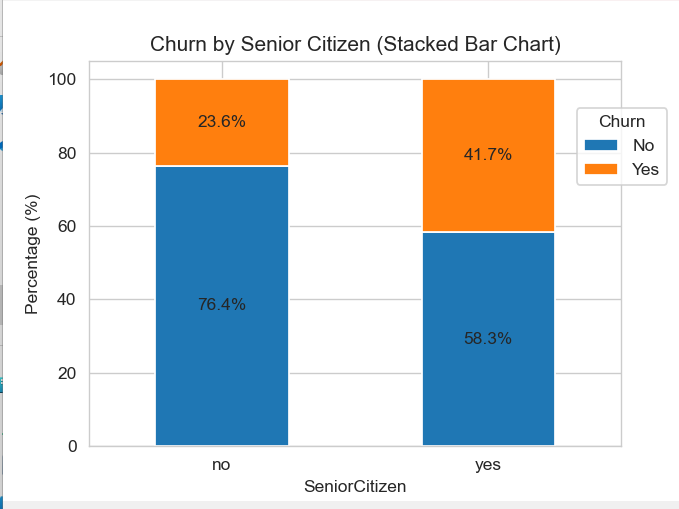
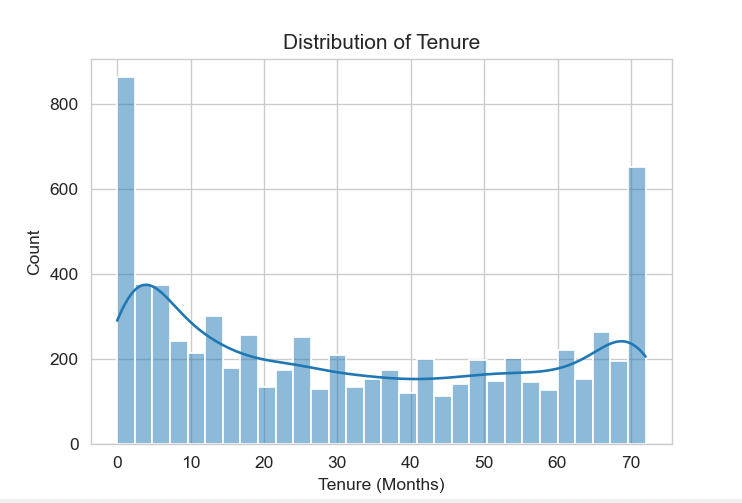
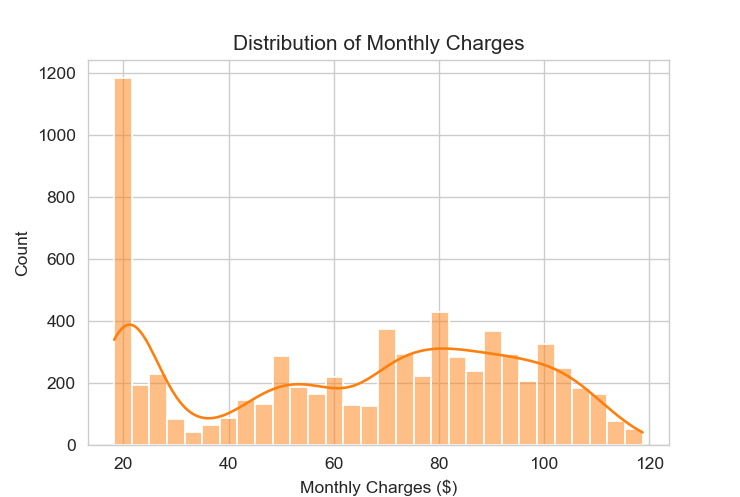
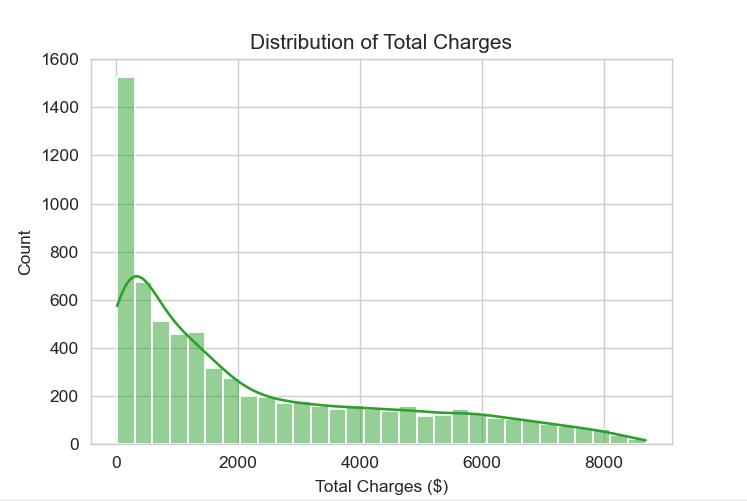
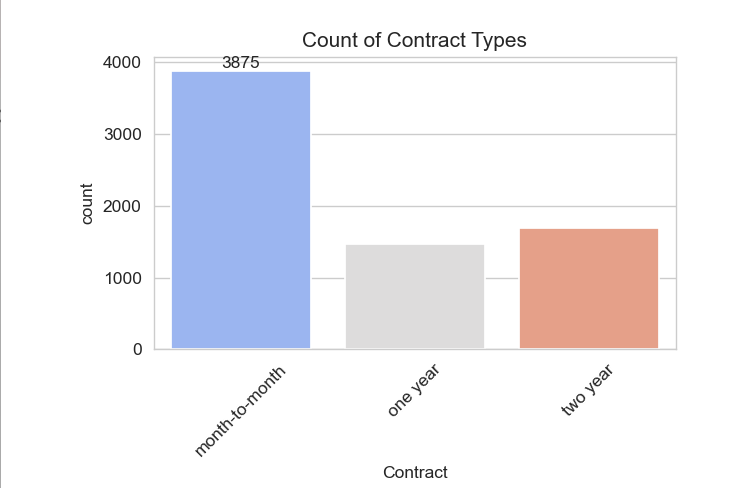
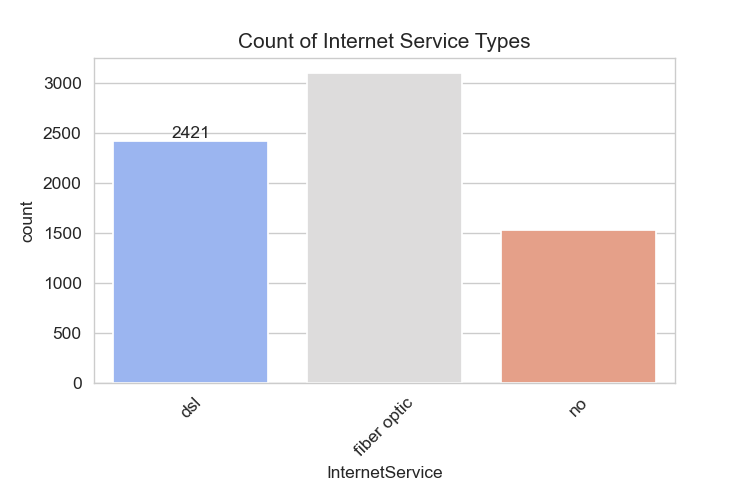
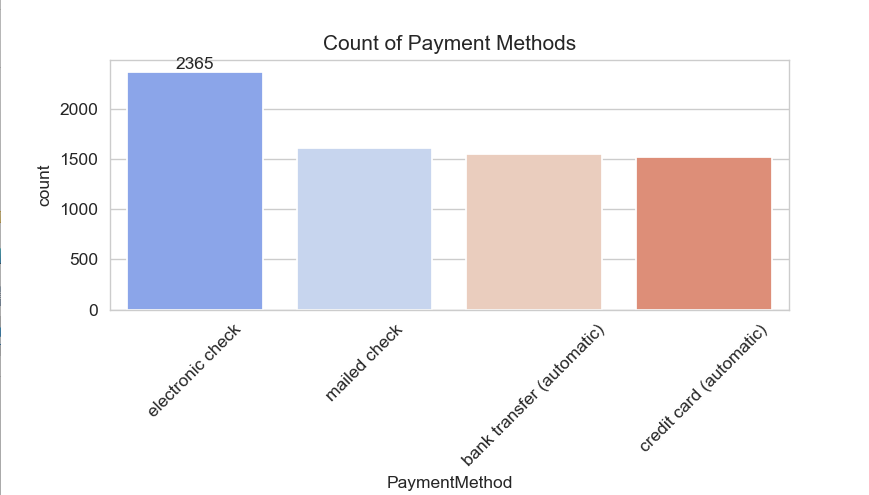
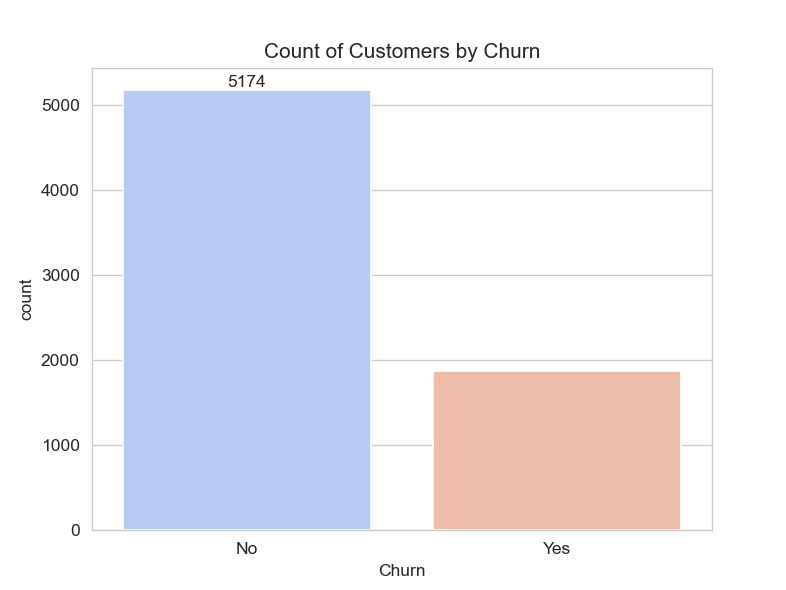
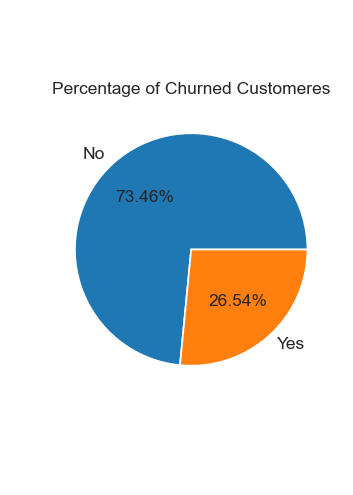
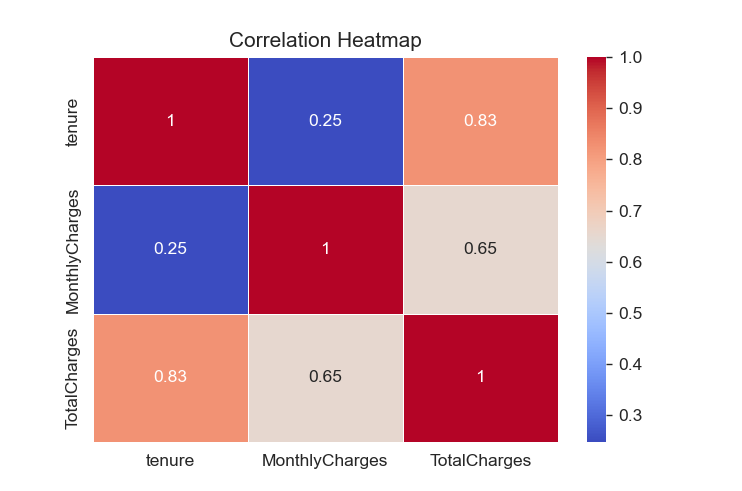
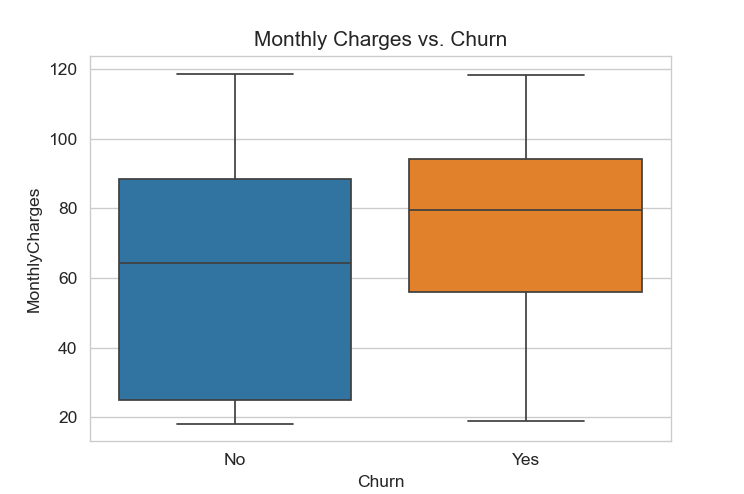
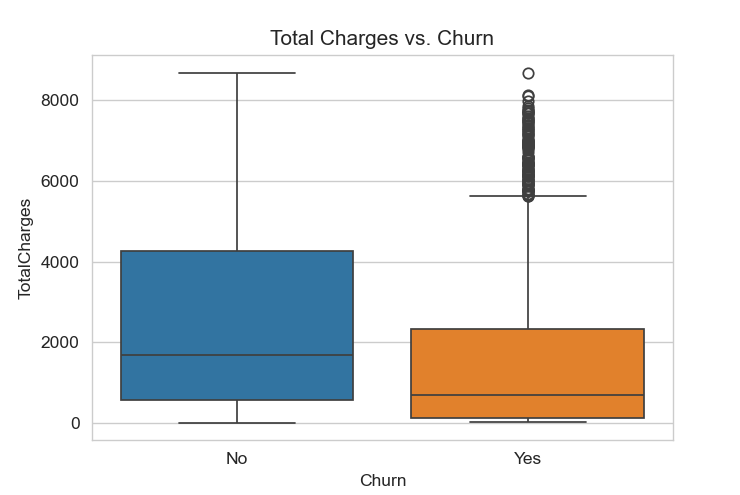
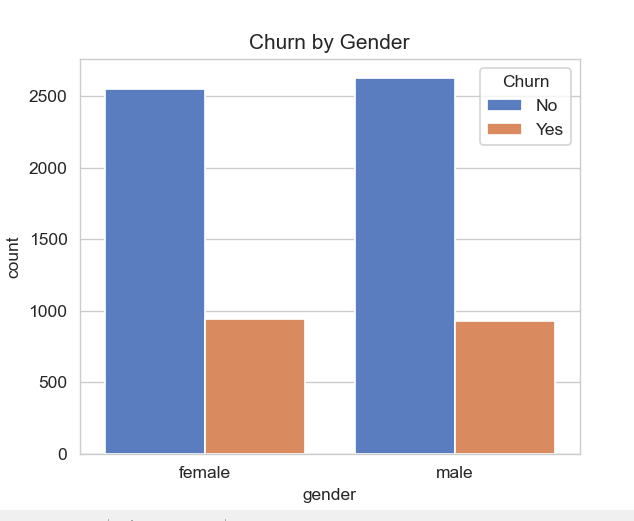
plt.figure(figsize = (3,4))

gb = df.groupby("Churn").agg({'Churn':"count"})

plt.pie(gb['Churn'], labels = gb.index, autopct = "%1.2f%%")

plt.title("Percentage of Churned Customeres", fontsize = 10)

plt.show()

**Visualizations**

**Conclusions**

In conclusion, the analysis shows that customers with shorter tenure, higher monthly charges, and month-to-month contracts are more likely to churn. Senior citizens also tend to leave at higher rates. To reduce churn, the company should focus on offering better deals for long-term contracts, lowering monthly charges where possible, and providing more support to senior customers. These steps can help improve customer retention and reduce the number of customers leaving.