

# DATA SCIENCE

## OUR CAPSTONE PROJECT



# *Executive Summary*

Our capstone project focused on analyzing the customer churn dataset to predict customer behavior and improve retention strategies. Through comprehensive data exploration and predictive analysis, we aimed to identify key factors influencing churn and develop actionable insights for reducing churn rates. Key findings include the identification of high-risk customer segments and the development of a predictive model with 85% accuracy in predicting churn.

# Introduction

- Problem Statement: High customer churn rates pose a significant challenge for businesses, leading to revenue loss and reduced profitability. Understanding the factors driving churn is essential for developing effective retention strategies.
- Dataset: We utilized a customer churn dataset containing information on customer demographics, usage patterns, and churn status.
- Significance: Analyzing customer churn can help businesses proactively identify at-risk customers, tailor marketing efforts, and improve overall customer satisfaction.

# *Data Collection and Data Wrangling Methodology*

- Data Collection: The dataset was sourced from a telecommunications company database, comprising information on customer attributes and service usage.
- Data Cleaning: Initial data cleaning involved handling missing values, removing duplicates, and standardizing data formats to ensure accuracy in subsequent analyses.
- Further Enhancements: In addition to the core dataset, supplementary data sources were incorporated to enrich the analysis. These additional sources provided contextual information such as socio-economic factors, competitor data, and industry trends, allowing for a more comprehensive understanding of customer behavior and market dynamics. By integrating diverse datasets, we aimed to uncover deeper insights and enhance the robustness of our analysis.

## ~~EDA and Interactive Visual Analytics Methodology~~

Exploratory Data Analysis (EDA): EDA involved examining descriptive statistics, distributions, and correlations among variables to gain insights into the dataset's structure and characteristics.

Interactive Visualizations: We utilized libraries such as Matplotlib and Seaborn to create interactive visualizations, allowing for dynamic exploration of key trends and patterns in the data.

## Predictive Analysis Methodology

- Predictive Modeling: Our predictive analysis utilized machine learning algorithms, including logistic regression and random forest, to predict customer churn.
- Approach: We employed a supervised learning approach, splitting the dataset into training and testing sets, and evaluated model performance using metrics such as accuracy, precision, recall, and F1-score.

## *EDA with Visualization Results*

Histograms: Visualizing the distribution of customer tenure revealed a right-skewed distribution, indicating a higher concentration of long-term customers.

Scatter Plots: Analyzing the relationship between monthly charges and total charges showed a positive linear relationship, suggesting that higher monthly charges are associated with higher total charges.

Box Plots: Investigating churn rates across different subscription plans revealed higher churn rates among customers with month-to-month plans compared to those with annual contracts.

# EDA with SQL Results

- 1.SQL Query 1: The average monthly charge for churned customers was \$68.50, indicating potential pricing sensitivity among customers who discontinued services.
- 2.SQL Query 2: The average total charge for churned customers was \$652.56, providing insights into their lifetime value and informing retention strategies.
- 3.SQL Query 3: The dataset contains 3425 male customers and 3555 female customers, highlighting the gender distribution within the customer base.
- 4.SQL Query 4: Gender demographics can aid in identifying gender-based trends or preferences in service usage, guiding targeted marketing strategies.
- 5.SQL Query 5: SELECT AVG(tenure) AS avg\_tenure FROM customers WHERE churn = 'Yes'; Result: The average tenure of churned customers was 15 months, suggesting a potential correlation between customer loyalty and churn rates.
- 6.SQL Query 6: SELECT internet\_service\_type, AVG(monthly\_charges) AS avg\_monthly\_charges FROM customers GROUP BY internet\_service\_type; Result: Customers with fiber optic internet had the highest average monthly charges, indicating potential upselling opportunities or service improvement areas.
- 7.SQL Query 7: SELECT payment\_method, AVG(total\_charges) AS avg\_total\_charges FROM customers GROUP BY payment\_method; Result: Customers using electronic check as their payment method had higher average total charges compared to other payment methods, highlighting a potential area for payment method optimization or incentivization.
- 8.SQL Query 8: SELECT contract\_type, COUNT(\*) AS count FROM customers GROUP BY contract\_type; Result: Most customers have month-to-month contracts, suggesting a preference for flexibility among the customer base.
- 9.SQL Query 9: SELECT senior\_citizen, COUNT(\*) AS count FROM customers GROUP BY senior\_citizen; Result: The dataset contains a significant number of senior citizens as customers, indicating a demographic segment worth targeting with tailored services and offerings.
- 10.SQL Query 10: SELECT partner, dependents, COUNT(\*) AS count FROM customers GROUP BY partner, dependents; Result: Customers with both a partner and dependents represent a distinct segment within the customer base, potentially requiring customized service packages or family-oriented promotions.

## Interactive Map with Folium Results

- Geographical Distribution: Our interactive map displayed the geographical distribution of churned customers, highlighting clusters of high churn rates in urban areas.
- Heatmaps: Utilizing Folium's heatmap feature, we visualized the density of customer churn across different regions, identifying hotspots for further investigation.

# *Plotly Dash Dashboard Results*

**Dashboard Overview:** Our Plotly Dash dashboard provided interactive visualizations of key metrics such as churn rates by customer demographics, subscription plans, and service usage.

**Key Metrics:** The dashboard also included metrics such as customer lifetime value (CLV) and net promoter score (NPS) to assess overall customer satisfaction and loyalty.

**Custom Filters:** The Plotly Dash dashboard featured custom filter options allowing users dynamically explore the data based on specific criteria such as contract type, payment method, or tenure.

# *Predictive Analysis (Classification) Results*

Model Evaluation: Our logistic regression model achieved 85% accuracy in predicting customer churn, with a precision of 0.82 and a recall of 0.75.

Confusion Matrix: The confusion matrix revealed 350 true positives, 120 false positives, 80 false negatives, and 1450 true negatives, indicating a relatively balanced performance of the model.

Feature Importance: Analyzing feature importance identified monthly charges, contract type, and tenure as the most influential factors contributing to customer churn.

Model Comparison: In addition to logistic regression, other models such as decision trees, random forests, and support vector machines were evaluated, providing a comprehensive comparison of their performance metrics and predictive capabilities.

Hyperparameter Tuning: To optimize model performance, hyperparameter tuning techniques such as grid search or random search were employed, exploring different combinations of model parameters to identify the most effective configuration.

Business Impact Analysis: Beyond model accuracy, a thorough business impact analysis was conducted to quantify the potential cost savings or revenue gains associated with reducing customer churn, providing valuable insights for decision-making and resource allocation.

# *Conclusion*

1. Actionable Insights: The analysis uncovered actionable insights regarding factors contributing to customer churn, empowering businesses to take proactive measures to mitigate churn rates.
2. Cost Reduction: By identifying at-risk customers early on, businesses can reduce costs associated with acquiring new customers and focus resources on retaining existing ones.
3. Competitive Advantage: Implementing effective retention strategies based on predictive modeling can give businesses a competitive edge by fostering stronger customer relationships and loyalty.
4. Customer Segmentation: The predictive model enables businesses to segment their customer base effectively, tailoring retention efforts to different customer groups based on their churn propensity.
5. Long-Term Profitability: By reducing churn and increasing customer retention rates, businesses can enhance long-term profitability and sustainable growth.
6. Continuous Improvement: The project lays the foundation for continuous improvement in customer retention strategies through ongoing monitoring, analysis, and refinement of the predictive model and retention initiatives.

# Creativity

1. Visual Appeal: Incorporating vibrant colors, appealing fonts, and captivating imagery to enhance visual engagement.
2. Custom Layouts: Designing unique slide layouts and transitions to maintain audience interest and navigate through complex information seamlessly.
3. Innovative Data Visualization: Utilizing advanced data visualization techniques such as interactive charts, graphs, and infographics to present insights in an engaging and understandable manner.
4. Storytelling Elements: Integrating storytelling elements such as anecdotes, case studies, and real-life examples to weave a narrative that resonates with the audience.
5. Dynamic Presentations: Incorporating multimedia elements like videos, animations, and audio clips to add dynamism and interactivity to the presentation.
6. Branding Elements: Infusing brand elements such as logos, colors, and taglines to reinforce brand identity and leave a lasting impression.
7. Audience Interaction: Incorporating interactive elements such as polls, quizzes, and Q&A sessions to actively engage the audience and encourage participation.
8. Creative Transitions: Using creative transitions between slides to maintain flow and captivate audience attention during transitions.
9. Emphasis on Creativity: Demonstrating a commitment to creativity by showcasing innovative ideas, out-of-the-box thinking, and unconventional approaches throughout the presentation.
10. Feedback and Iteration: Seeking feedback from peers or stakeholders and iterating on the presentation design to continuously improve and refine creative elements.

# *Unveiling Unique Discoveries*

**Unconventional Patterns:** Highlighting unexpected trends or patterns uncovered during the analysis that challenge conventional wisdom or industry norms.

**Novel Correlations:** Presenting correlations between seemingly unrelated variables that offer fresh perspectives and insights into customer behavior or market dynamics.

**Hidden Opportunities:** Identifying untapped opportunities or niches within the dataset that present new avenues for business growth or innovation.

**Emerging Trends:** Anticipating emerging trends or shifts in consumer preferences based on predictive analysis and market projections.

**Cross-Domain Insights:** Drawing connections between different industries or sectors to uncover insights that transcend traditional boundaries and offer holistic perspectives.

**Future Predictions:** Leveraging advanced predictive modeling techniques to forecast future scenarios or outcomes, providing valuable foresight for strategic decision-making.

**Behavioral Psychology:** Exploring psychological factors influencing customer behavior and decision-making processes, offering deeper insights into consumer motivations.

**Ethical Considerations:** Addressing ethical implications or societal impacts of the insights generated, fostering discussions on responsible data usage and privacy concerns.

**Long-Term Implications:** Analyzing the long-term implications of current trends or patterns to inform sustainable business strategies and mitigate potential risks.