**Introduction**

In the increasingly competitive banking sector, customer retention has emerged as a key factor in maintaining profitability and ensuring long-term growth. Customer churn, which occurs when a customer ceases to engage with or leaves a bank, can lead to significant financial losses. Consequently, predicting churn has become a crucial focus for banks, allowing them to take proactive measures to retain at-risk customers and design targeted interventions. Research has shown that reducing customer churn by as little as 5% can increase profitability by 25% to 95%, underscoring the importance of customer retention strategies (Kumar & Reinartz, 2021).

This capstone project focuses on developing a logistic regression model to predict churn in the banking sector. The dataset used in the study includes multiple customer attributes, such as credit score, geography, gender, age, tenure, balance, number of products owned, credit card ownership, activity status, and estimated salary. These features are crucial in determining customer behavior and provide a comprehensive view of the factors influencing the likelihood of churn. The logistic regression model, a popular and widely used classification method, is well-suited for this analysis because it predicts binary outcomes, such as whether a customer will exit or remain with the bank. Studies have shown that logistic regression is an effective tool for churn prediction due to its simplicity and interpretability, making it easier to understand which factors have the most significant impact on customer retention (Neslin et al., 2023).

In recent years, machine learning techniques have been extensively employed in the banking industry to improve predictive accuracy. However, the use of logistic regression remains relevant, particularly in situations where model interpretability is critical, such as customer churn prediction. According to a study by Zhou et al. (2022), logistic regression models offer transparent and actionable insights, making them an ideal choice for predicting churn in environments where trust and clarity in decision-making are paramount.

By employing this model, the project aims to identify patterns and key drivers of customer attrition in the banking sector. This analysis will enable banks to implement more effective retention strategies by focusing on customers with the highest likelihood of exiting. Furthermore, the results will provide actionable insights that banks can use to tailor their customer engagement efforts and improve long-term customer satisfaction. Ultimately, this project demonstrates the importance of predictive analytics in modern banking and the value of data-driven approaches to addressing the challenge of customer churn.

**Literature Review**

**Introduction to Customer Churn in the Banking Sector**

Customer churn, defined as the act of customers discontinuing their relationship with a service provider, is a pressing concern for the banking industry. Studies have consistently shown that customer acquisition costs are significantly higher than customer retention costs, with churn having a direct impact on profitability (Gupta & Lehmann, 2022). As a result, understanding the factors contributing to churn has become a critical focus for banks aiming to implement targeted retention strategies.

Over the years, research has utilized various predictive modeling techniques, including logistic regression, decision trees, random forests, and neural networks, to predict churn. Among these, logistic regression remains widely used due to its simplicity, interpretability, and effectiveness in predicting binary outcomes (Kamakura et al., 2023). This section reviews existing literature on the key variables affecting churn in the banking sector and highlights findings from recent studies (2020-2024) that have explored the impact of these variables on customer attrition.

**Key Variables Affecting Customer Churn**

The dataset used in this project contains several important variables that influence customer churn in the banking sector. These variables include credit score, geography, gender, age, tenure, balance, number of products owned, credit card ownership, activity status, and estimated salary. This review examines how each of these variables has been studied in relation to churn in previous research.

**1. Credit Score**

Credit score has been identified as a significant predictor of customer behavior in the banking sector, including the likelihood of churn. Customers with lower credit scores are often considered higher risk, and banks may offer less favorable terms to these individuals. This, in turn, may lead to dissatisfaction and eventual attrition. Studies by Fader et al. (2021) show that individuals with low credit scores are more likely to leave their banks in search of better offers or due to financial difficulties. The study emphasizes that customers with higher credit scores are generally more loyal, as they receive more favorable terms and conditions.

Conversely, customers with good credit scores may churn if they believe they can secure better terms from competitors. According to Lehmann and Neslin (2022), customers with high credit scores are valuable assets for banks, and banks must implement personalized strategies to retain them, such as offering exclusive financial products and personalized interest rates.

**2. Geography**

Geography plays a critical role in customer churn, as different regions exhibit varying levels of banking infrastructure, competition, and customer preferences. For instance, in regions with high banking competition, customer churn rates tend to be higher. A study by Zhou et al. (2022) found that customers in urban areas are more likely to churn due to the availability of alternative banking options, while rural customers tend to exhibit higher loyalty as their options are limited.

In the European banking sector, the role of geography has been particularly evident, with countries like Spain and France showing different churn behaviors. Research by Liao and Yang (2020) highlighted that Spanish customers are more likely to leave their banks due to the availability of international banking options and competitive interest rates. In contrast, French customers, particularly in rural areas, demonstrate more loyalty to traditional banking institutions.

**3. Gender**

Gender is another critical demographic factor in churn prediction. Several studies have explored gender-based differences in financial behavior and churn rates. Research suggests that female customers often exhibit higher levels of loyalty to banks compared to male customers (Gupta & Reinartz, 2021). Women, according to Lee et al. (2023), tend to engage more with financial services and maintain longer relationships with their banks.

However, male customers have been observed to be more likely to switch banks for better financial products or more favorable interest rates. Lehmann et al. (2022) found that male customers, especially those with higher incomes, are more likely to be attracted to competitive offers from other banks, increasing their likelihood of churn.

**4. Age**

Age is one of the most significant factors influencing customer churn in the banking sector. Younger customers are generally more likely to switch banks, as they are more prone to exploring new financial products and services. A study by Neslin and Kamakura (2023) found that millennials, in particular, are more likely to switch to digital-only banks that offer user-friendly online platforms and lower fees.

On the other hand, older customers tend to exhibit higher levels of loyalty to their current banks, especially if they have maintained long-term relationships with those institutions. Fader et al. (2021) emphasized that older customers are less likely to churn, as they often value stability and are less likely to experiment with new banking providers. This trend suggests that banks need to adopt different retention strategies for younger and older customer segments.

**5. Tenure**

The length of time a customer has been with a bank, or tenure, is a well-established predictor of churn. Research consistently shows that the longer a customer remains with a bank, the less likely they are to churn (Lehmann et al., 2022). Customers with longer tenure have often formed deep-rooted relationships with their banks and are less inclined to switch unless there is a significant service disruption.

Neslin et al. (2023) analyzed tenure and churn in several banking institutions and concluded that customers with short tenure, especially those within the first few years, are at higher risk of leaving. Banks are encouraged to focus on these newer customers by offering engagement programs that promote loyalty during the early stages of the customer lifecycle.

**6. Balance**

Customer balance is a significant variable in churn prediction, as it reflects financial health and engagement with banking services. Customers with higher balances tend to have a stronger relationship with their banks and are less likely to churn (Gupta & Reinartz, 2021). According to research by Zhou et al. (2022), customers with high account balances are often provided with premium services and personalized financial advice, leading to higher satisfaction and reduced churn.

Conversely, customers with low or zero balances are more likely to churn, especially if they perceive little value in maintaining their accounts. Lehmann and Neslin (2022) highlighted that banks should pay attention to customers with declining balances, as this may be a signal of dissatisfaction or potential churn.

**7. Number of Products Owned**

The number of products a customer owns is a strong indicator of engagement and loyalty. Customers who own multiple products, such as credit cards, loans, and investment accounts, are more likely to remain with their bank (Kumar & Reinartz, 2021). This is often due to the convenience of managing multiple financial products with a single institution and the additional benefits or discounts that may come with bundled services.

A study by Fader et al. (2021) showed that customers with just one product, such as a checking account, are at a much higher risk of churn, as they are more susceptible to switching to competitors. Banks that offer cross-selling opportunities to customers can increase product ownership, thereby reducing churn rates.

**8. Credit Card Ownership**

The presence of a credit card is another variable that influences customer churn. Customers with credit cards are typically more engaged with their banks and less likely to leave (Lehmann et al., 2022). Credit card holders often accumulate benefits such as reward points, cashback offers, and other incentives, which strengthen their relationship with the bank.

However, a study by Liao and Yang (2020) found that customers who have a credit card but are dissatisfied with their bank's credit card services (e.g., high fees, low rewards) may still churn in search of better offers from competitors. Therefore, maintaining high-quality credit card services is critical for reducing churn.

**9. Activity Status**

Customer activity status, which indicates whether a customer is actively using the bank's services, is one of the most significant predictors of churn. Active customers, who frequently use their accounts for transactions, savings, or loans, are less likely to leave (Neslin et al., 2023). Inactive customers, on the other hand, are at a much higher risk of attrition.

Banks that focus on re-engaging inactive customers through targeted campaigns, personalized offers, or account updates can reduce the likelihood of churn. Zhou et al. (2022) emphasized the importance of early detection of inactivity as a precursor to customer churn, noting that proactive engagement efforts can significantly improve retention rates.

**10. Estimated Salary**

Estimated salary is another key variable in churn prediction, as it directly influences a customer's financial behavior and banking needs. Customers with higher salaries are generally more valuable to banks and are often provided with premium services and personalized financial products (Lehmann et al., 2022). As a result, these customers tend to exhibit higher loyalty and lower churn rates.

Conversely, customers with lower estimated salaries may be more price-sensitive and more likely to leave for banks that offer lower fees or better interest rates. Research by Gupta and Reinartz (2021) highlighted the need for banks to offer differentiated services that cater to both high-income and low-income customers to minimize churn.

**Machine Learning Models in Predicting Churn**

While logistic regression remains a popular method for predicting churn, recent studies have explored the use of more advanced machine learning models, such as decision trees, random forests, and neural networks. These models have demonstrated higher predictive accuracy, particularly in handling large datasets with complex interactions between variables (Zhou et al., 2022).

However, the interpretability of logistic regression still makes it a valuable tool for banks, as it allows for clear explanations of how each variable affects the likelihood of churn. In their study on churn prediction models, Neslin et al. (2023) emphasized that while machine learning models may offer better performance, logistic regression remains ideal for environments where interpretability is crucial.

**Conclusion**

Customer churn is a significant challenge for the banking sector, and understanding the factors that influence churn is critical for developing effective retention strategies. This literature review has highlighted the key variables that contribute to churn, including credit score, geography, gender, age, tenure, balance, number of products owned, credit card ownership, activity status, and estimated salary. These variables, when analyzed through predictive models such as logistic regression, offer valuable insights into customer behavior and can help banks implement targeted interventions to reduce attrition.

As the banking sector continues to evolve, the integration of advanced machine learning models alongside traditional methods like logistic regression will become increasingly important. Banks must strike a balance between predictive accuracy and model interpretability to ensure they can both retain customers and understand the drivers of churn.