### Literature Review

### In the past decade, data analysis has become increasingly important in the banking industry and is widely used in areas such as fraud detection, customer behavior analysis, credit risk assessment and market trend prediction.The study by Nobanee et al. (2021) review the existing literature on big data applications in banking using a bibliometric analysis approach. They point out that applications of big data in the banking sector are growing rapidly and highlight its impact on areas such as risk management, customer relationship management, and investment banking. [银行业的大数据应用：文献计量分析方法 - Haitham Nobanee、Mehroz Nida Dilshad、Mona Al Dhanhani、Maitha Al Neyadi、Sultan Al Qubaisi、Saeed Al Shamsi，2021 年 (sagepub.com)](https://journals.sagepub.com/doi/10.1177/21582440211067234#bibr43-21582440211067234)

### The article by N. Sun et al. introduces iCARE, a framework for big data-based banking customer analytics. Practical examples of the iCARE framework show that effective data integration and analysis can significantly improve the personalization of banking products and services, thereby bringing commercial advantages to banks in a highly competitive market. [iCARE: A framework for big data-based banking customer analytics | IBM Journals & Magazine | IEEE Xplore](https://ieeexplore.ieee.org/abstract/document/6964895#full-text-header)

### Srivastava and Gopalkrishnan describe how Indian banks are using big data to analyze transaction behavior to segment customers for targeted marketing strategies, such as cross-selling financial products based on individual spending behavior. This article provides us with inspiration on how to apply big data to improve customer satisfaction and bank profitability. [大数据分析对银行业的影响：印度银行的学习 - ScienceDirect (oclc.org)](https://www-sciencedirect-com.bris.idm.oclc.org/science/article/pii/S1877050915005992)

### Research by Delgosha et al.(2020) revealed that the most important applications of big data in banks are fraud detection and credit risk analysis. [Elucidation of big data analytics in banking: a four-stage Delphi study | Emerald Insight (oclc.org)](https://www-emerald-com.bris.idm.oclc.org/insight/content/doi/10.1108/JEIM-03-2019-0097/full/html)

### In Chang and Kim's research, they showed how to use big data technology, combined with traditional demographic information and customers' transaction behavior data, to effectively segment bank customers through machine learning methods such as self-organizing mapping (SOM). This enables banks to precisely tailor marketing strategies and products to the specific needs of different customer groups. [基于银行大数据的客户细分方案 -数字内容学会杂志 |韩国科学 (koreascience.kr)](https://koreascience.kr/article/JAKO201808962641880.page)

### The insights gained from these studies illuminate several paths we can take to analyze our data sets. By applying visual analytics, we can discover hidden patterns and trends in transaction data, similar to the visual techniques proposed by Chang and Kim. The customer behavior analysis demonstrated by Srivastava and Gopalkrishnan will enable us to effectively segment the customer base, thus aiding in personalized marketing and product development. Additionally, inspired by the approach used by Delgosha et al., we will focus on enhancing our fraud detection capabilities, leveraging predictive models to identify and prevent potentially fraudulent activities.

### Methodology

### EDA

### Exploratory data analysis (EDA) is a basic process of data analysis that examines the main characteristics of a data set in a visual and statistic way. It summarizes key characteristics of data through charts, plots, and statistics to identify patterns, anomalies, and relationships to gain insight into the underlying structure and quality of the data. This key step helps us understand the structure and quality of the data to guide further analysis.

### . Data Visualization with Tableau

**Tableau works great for visualizing our datasets because it excels at handling large data volumes and has an intuitive, interactive interface that makes complex and large data patterns easy to access and understand.**

### User Analysis Techniques

### K-means clustering: Segmenting customers based on transaction amount and frequency by dividing them into a specified number of clusters, which helps in classifying customers into different groups based on their transaction behavior.

### Hierarchical clustering: Construct a tree diagram that illustrates how customers are linked based on their transaction data, enabling visualization of customer relationships across multiple similar levels.

### RFM analysis: Divide customers into segments based on three factors: when they last purchased (recency), how often they purchased (frequency), and how much the customer spent (currency).

### RFM-based CLV: Combines RFM models with predictive analytics to estimate future profitability from customers and calculate customer lifetime value based on past transaction patterns.

### Fraud Detection Methods

#### Isolation Forest

#### **Detect anomalous transaction behavior by isolating outliers in the data, which can effectively identify data points that are significantly different from the majority of data points, thereby potentially indicating fraudulent transactions.**

#### Neural Networks

Neural networks can detect fraud by identifying typical spending behavior based on large amounts of transaction data and flagging transactions that deviate significantly from these patterns.

#### Decision Trees and Random Forest

Use decision tree algorithms to provide intuitive decision paths based on data attributes. Random Forest improves this by creating an ensemble of decision trees to increase the accuracy and robustness of predicting and classifying data, helping to differentiate between legitimate and fraudulent transactions.

**Results**

We used Tableau visualization tools to analyze bank transaction data, revealing trends in consumer behavior, transfer patterns, and customer segmentation based on deposit levels, providing suggestions to improve bank operations and enhance customer experience.

**Time Analysis:**

The first data set shows that weekend consumption is significantly higher than weekdays, especially in the first three weekends of December, when consumption reaches an annual high, which may be related to the holiday shopping season. The second data set shows that consumption reaches significant peaks on the first and last days of each month, possibly due to cyclical large transactions such as loan repayments and payroll payments. Non-transfer consumption is mainly concentrated between 8:00 am and 20:00 pm. The peak consumption on weekdays is 9:00 and 17:00, and the peak consumption on Saturdays and Sundays is 10:00.

**Merchant Analysis:**

We divided merchants into seven different types. We found that people spend more on coffee to work efficiently during weekdays, while weekends see more spending at bars and restaurants as people relax and catch up with family and friends.

**Transfer Analysis:**

We can also see that the total amount of transfers on weekends is about 1.5 times that of weekdays, but the frequency is ten times higher, indicating that people may make large transactions on weekdays and many small transfers on weekends. This may indicate that people may make large work transactions during the week and split the bill with family and friends while hanging out on the weekends. This may suggest that people split the bill with their family and friends while hanging out on weekends. Therefore, we can offer our customers with a convenient split payment service.

**Customer Segmentation Analysis:**

Through classification and visual analysis of bank customers with different deposit levels, we found that the spending of low-deposit customers is mainly concentrated in Halifax and LBG, which may be mainly loan repayments or basic financial services, indicating that these customers have low spending power. There may be higher financial stress. Therefore, banks can consider providing more flexible repayment plans or low-interest financial products to such customers to help them better manage financial stress. For high-deposit and high-asset customers, their main income comes from financial services, advanced medical care and education industries, showing high economic activity and spending power. Banks can provide these customers with high-end financial services, investment consulting and customized financial solutions to meet their needs for complex financial products and services.

**Discussion**

In our dataset, K-means clustering and hierarchical clustering methods were not chosen, mainly because the requirements of these two methods on the dataset did not exactly match the parameters we had. K-means clustering requires rich feature dimensions to effectively distinguish different customer groups, but our data mainly includes basic transaction information, such as transaction amount and date, and lacks more dimensions of customer behavior characteristics. In addition, K-means is more sensitive to outliers, and extreme transaction values common in financial data may lead to unsatisfactory clustering results. Hierarchical clustering provides a detailed view of the data structure, but its computational complexity is high, time-consuming for large-scale data processing, and it is also sensitive to noise and outliers. These factors make these two methods potentially unsuitable for direct application to our simple dimensional data set without optimization to accommodate the specificities of financial data.

**Further Improvement**

In order to analyze user behavior more effectively and improve service quality, banks can consider legally collecting the following additional information: transaction type (such as deposits, withdrawals, payments, etc.), customer's age, gender, occupation and education level. In addition, it is useful to include residential address information to analyze geographic spending patterns, transaction frequency and size, and the device through which transactions were completed (mobile banking, online banking, or traditional methods). Understanding a customer's credit history and the specific location of transactions can also help banks assess credit risk and prevent fraud. When collecting this data, banks need to ensure compliance with relevant data protection regulations, ensure transparency in processing data and obtain customer consent.