Intro

今天我们将展示劳埃德银行提供的 DSMP 问题 C 的数据分析结果。我们得到的第一个数据集包括转账金额、交易双方账户和转账日期等信息。第二个数据集还包括交易的时间戳和余额。为了避免潜在的隐私侵犯和其他道德问题，我们获得的数据集是在真实数据基础上训练出来的虚构数据。不过，随后描述的分析方法和可视化效果可应用于真实数据集。

Good Morning, thank you for taking the time to join us today. My name is Shuyao Yi, and I am here on behalf of Group 25. Today we will show the results of data analysis provided by Lloyds Bank. The first dataset we have obtained includes information on the amount, the accounts of both parties and the date of transaction. The second dataset also includes the timestamp of transaction and balance of the customers To facilitate subsequent analyses, we grouped the merchants in the first dataset into six types and those in the second dataset into ten types. In order to avoid potential privacy violations and other ethical issues, the datasets we obtained are fictitious data trained on real data. However, the analysis methods and visualizations we used can be applied to the real dataset.

文献综述

通过文献综述，我们了解到大数据在银行业的应用主要包括风险管理、客户关系管理和提高营销战略水平方面。通过分析银行日志和信用卡交易数据，我们可以绘制用户画像，来提供个性化金融服务；通过评估客户价值，我们可以识别高价值客户和制定针对性的营销策略；大数据分析还可提升银行对异常交易的识别和欺诈防控能力。因此，我们确定了针对我们数据集的三个研究方向，分别是评估客户价值、研究客户消费习惯和客户安全保障。

Through the literature review, we have learnt that the applications of big data in the banking industry mainly include risk management, customer relationship management and marketing strategy development. By analysing bank logs and credit card transaction data, we can draw user profiles to provide personalised financial services. By assessing customer value, we can identify high-value customers and formulate targeted marketing strategies. By analysing abnormal transaction data, we can also identify fraud. Therefore, we chose three research directions for our data set, which are assessing customer value, studying customer spending habits, and ensuring customer account security.

在我们的数据分析中，我们首先运用了 Tableau 这一可视化工具，对我们的数据集展开了初步分析。基于两个数据集的不同特点，我们进行了不同方向的分析。在客户价值方面，我们筛选了第二个数据集中总收支绝对值大于200k的商家，发现低存款客户消费主要集中在哈利法克斯和 LBG，可能是因为贷款偿还。相比之下，高存款和高收入客户的主要收入来自金融服务、先进医疗保健和教育行业。

在客户消费习惯方面，我们进行了时间分析。第一个数据集显示周末消费高于工作日，特别是在 12 月的前三个周末，可能因为假日购物季。第二个数据集则显示，每个月的第一天和最后一天消费和收入都达到高峰，可能是由于周期性大额交易，如贷款偿还和工资支付。工作日消费高峰在上午 9:00 和下午 5:00，周末在上午 10:00。

在商户分析方面，我们将第一个数据集中的商户分为六种不同类型。我们发现人们在工作日更多地在咖啡馆消费以提高工作效率，而在周末，他们更多地在酒吧和餐馆消费，放松身心。

最后，在交易分析方面，周末转账总金额约为工作日的 1.5 倍，但转账频率要高出十倍。这表明人们可能在工作日进行大额交易，而在周末与家人和朋友一起外出时进行许多小额转账。而在第二个数据集中，我们将不同收入的客户进行分类，以5000镑的存款作为阈值，低收入群体主要转账为银行间贷款偿还、日常必需品支出，高收入群体则主要来自医院、护理和金融。

We used Tableau to initially analyse our datasets. Based on the different characteristics of the two datasets, we conducted analyses in different directions.

In terms of customer value, we found in the second dataset that low-deposit customers' spending is mainly concentrated in Halifax and LBG, probably due to loan repayments. In contrast, high deposit and high income customers derive their income mainly from the financial, healthcare and education sectors.

In terms of customer spending habits, we conducted a temporal analysis. The first dataset shows higher spending on weekends than on weekdays, especially during the first three weekends of December, possibly because of the holiday shopping season. The second dataset, shows that consumption and income peak on the first and last days of the month, possibly due to cyclical large transactions such as loan repayments and salary payments. Consumption peaks at 9:00 a.m. and 5:00 p.m. on weekdays and at 10:00 a.m. on weekends.

For merchant analysis, we found that people spend more in cafes on weekdays to be productive, while on weekends they spend more in bars and restaurants to relax.

In transaction analysis, the ratio of average transfer frequency to total amount on weekends indicates a higher occurrence of small transfers, while weekdays are more prone to large transfers.

And in the second dataset, we found the lower income groups mainly transferring money for loan repayments and everyday essentials, and the income for higher groups are mainly from hospitals, care and finance.

基于前面的可视化分析，我们发现账户普遍存在一些转账时间，频率和金额的规律，因此我们可以通过检测这三个方面的异常来进行欺诈分析。之后会由其他组员详细介绍，接下来有情别的组员介绍higher level analysis。

Based on the previous visual analysis, we found that there are some general patterns of transfer time, frequency and amount, so we can detect the anomalies in these three aspects for fraud analysis, which will be introduced in detail by other group members later. Next, I would like to invite other group members to introduce higher level analysis.