APPLICATIONS OF
MODELLING AND SIMULATION
http://arqiipubl.com/ams eISSN 2600-8084 VOL 5, 2021, 184-190

Simulating Data Stories of Clients' Credit Card Default

Siti Salwa Salleh*, Nor Ayunie Mohamed and Nur Ainin Sofea Arman Shah

Fakulti Sains Komputer dan Matematik, Universiti Teknologi MARA, Cawangan Negeri Sembilan, Kampus Seremban, Negeri Sembilan, Malaysia

*Corresponding author: ssalwa@uitm.edu.my

Submitted 01 September 2021, Revised 09 October 2021, Accepted 12 October 2021.

Abstract: Clients who have problems paying with their credit cards need to be identified. However, data analysis is rather difficult due to the numerical nature of raw data. In the digital era, it is necessary to deploy dashboards and visual items to assist users in obtaining information more quickly. Typically, the dashboard comprises relevant charts and graphics. In this study, a Sankey chart is recommended. The purpose of this project is to simulate the construction of a dashboard that supports users at credit agencies or financial institutions in detecting credit card payment defaults. User requirements, data pre-processing, exploratory data analysis (EDA), data treatment, modelling, and assessment are all part of the development process. It is recommended to use a Sankey chart to examine patterns based on delayed monthly patterns. The result of this study shows that the largest percentage (71%) of the clients who do not have default payments in the coming month are those who are using revolving credit, paying duly, or those who do not consume the credit card. The usability assessment yielded a 4.0 mean on a five-point Likert scale which indicates that respondents agree that the dashboard is usable. In future development, a prediction model will be created and included into the default payment indicator.

Keywords: Credit card; Data storytelling; Dashboard; Payment default; Sankey chart.

1. INTRODUCTION

Data visualisation is the process of presenting data in order to assist the user's comprehension. It can be as simple as a two-dimensional graph or as complex as a three-dimensional graph. It is supposed to be useful as it simplifies information, highlights the most essential data, and conveys critical stories fast [1]. In contrast, a dashboard is a snapshot of a large set of information. It simply shows a portion of it, yet that portion is representative of the whole set of information [2]. The purpose is to provide an overall view of the current state of conditions to the user without presenting too much detail. A dashboard must be readable and understandable for non-technical users to ensure that they receive useful insight and make timely decisions. The type of chart used facilitates comprehension to which choosing the incorrect chart type may confuse consumers or result in data misinterpretation.

In general, digital dashboards are intended to be intuitive and simple-to-use front ends for monitoring, analysing, and optimising key business processes by empowering users at all hierarchical levels to make better decisions [3-6]. Digital dashboards are useful if they support their users in accomplishing goals [7]. Managers and executives in business require tools to aid them in decision-making process. In risk management and financial risk control, identifying potential credit card default clients and decreasing credit card defaults are major areas to focus on for financial institutions [2]. One of the most significant issues is identifying non-risky clients, but it is also the most difficult job for banks as the numerical aspect of raw data makes it difficult to understand and appreciate the data story beneath it [8]. Payment behaviour is variable and so difficult to observe with naked eyes. Numerous studies have been conducted to predict future default payments by clients using machine learning technique [3, 9-15]. However, these predictions are better presented in graphical form [16].

With credit card usage, clients will be able to make transactions and purchases at their convenience, but this may result in debt that must be paid off each month. When a client falls behind on their payments, a credit card default happens. Defaulting on a credit card has far-reaching consequences that affect not just the credit card company's relationship with the client, but also the payment history in general and the client's ability to acquire credit from other sources. To predict the client's default payment for the following month, credit card issuers have to study the client's background. Prior to that, payment behaviour for each month must be studied first. By analysing default payment data, banks and financial institutions will be able to forecast whether or not their clients will fail on paying their credit cards in the next month if the prediction is true. The procedure is carried out at least monthly as a consequence of the large-scale data acquired from the source.

To look into this issue, a dashboard is one of the approaches to take a data story and communicate it using the appropriate graphical form. This study explores the design of the dashboard and the visuals that are important to be included in it. To

demonstrate the findings, this study proposes a prototype that has been developed. In order to achieve the goal, the study's objectives are as follows: (i) to identify the client's profile, credit limit, and repayment behaviour, (ii) to establish a link between credit limit and the frequency of the delay pattern, and (iii) to depict the flow of a story involving four key nodes: basic demography, credit limit, repayment status for each given month, and the delay pattern. This paper is organised in the following format to present the work completed. It starts with an introduction that covers the problem's definition and objectives. Then, in Section 2, a literature review is presented, followed with Section 3 that presents the methodology. Section 4 discusses results and findings, and in the final section, the concluding remarks and future work are provided.

2. LITERATURE REVIEW

Data visualisation and dashboards are frequently used in conjunction. Both are effective in conveying specific information, and they work well when used together. Dashboards can be designed in various ways. There is no single correct or incorrect way – it all depends on the requirements for the dashboard [4, 7]. For instance, dashboards used to monitor operations must be built differently than those used to facilitate strategic decision-making or data analysis, and yet their design still depends on the user requirements [2]. The design can also vary depending on the usage scenarios, either for "pull" or "push" purposes [2]. A dashboard built to serve the pull scenario should include more options for exploring the data, filtering, and searching, investigating the causes of the data, and so on. When a user wants to get a certain piece of information, he or she uses the dashboard to do so. On the other hand, the dashboard in the push scenario must be designed in such a way that important information is pushed to the user. The dashboard must capture and enhance the user's attention. The purpose determines whether a dashboard is better suited for a push or pull scenario. It should literally assist users in comprehending the significance of the data by presenting it in a logical manner and allowing the user to select the level of data information needed [17].

The most common and instructive visualisation objects used to construct dashboards are bar graphs, line graphs, and bullet graphs [18]. According to another study, the most effective visualisation objects for a quick comparison are bar graphs, line graphs, and bullet graphs [19]. Dashboards should provide easy-to-use features and perception [17]. Users may not understand or be able to see any pattern if information is given in unexpected ways, as the human mind is trained to specific norms of display [4].

Apart from the visualisation objects mentioned above, a Sankey chart also provides a good data story. It can represent probability distributions at various levels of detail and there is a researcher who studied the potential of the Sankey chart to improve the visual flow of information in a machine learning system [20]. In addition, the Sankey chart is able to focus on a single aspect that needs to be highlighted at multiple viewing levels. Lamer *et al.* [21] used Sankey chart to show the proof of concept of a population's trajectory following an event. To assess the representation, two case studies in populations from Lille University Hospital's anaesthetic data archive were employed. The Sankey chart has been designed to highlight typical care paths for patients who are admitted to an intensive care unit. Based on the experiment outcome, they concluded that this initial work is well received by end-users.

In terms of design approach, Janes *et al.* [7] applied the GQM + Strategies measurement model as a basis for developing a dashboard. They developed a dashboard that follows a "push" usage scenario and evaluated it. Based on the acceptance test, they concluded that the dashboard is perceived as useful and easy to use. Apart from designing based on business goals, the development also involves management and experienced collaborators. The visualisation objects adopted are crucial. They have a huge impact on the acceptance of the dashboard [7, 22]. Similarly, Mazumdar *et al.* [23] described that the visualisation of dashboards is based on features of user preferences, usage history, the current task, the scale of datasets and types of data.

Many of the related studies on credit card defaults focused on prediction. Li *et al*. [24] used a weighted SVM method with differential privacy to create a model to predict credit card default. The technique is claimed to be able to reliably anticipate a client's default while preserving personal information. Another work presented an artificial neural network-based model that predicts clients who are likely to default on payment in the next month [25]. Ma [10] used the XGBoost model to analyse the data to predict a client's readiness to repay a credit card debt with additional financial characteristics derived from the original variables. Sharma and Mehra [8] also conducted a study to determine ways to identify risky and non-risky by examining the results of several algorithms implemented.

3. METHODOLOGY

There were five phases to this study: understanding the nature of the business, data understanding, data preparation and treatment, modelling, and assessment. Figure 1 depicts the overall steps of the study. The dataset used for this study is obtained from the UCI Machine Learning Repository [26, 27]. The dataset contains credit card payback between April and September 2005 of credit card issuers in Taiwan. The total number of records is 30,000.

Step 1 began with an effort to understand the nature of the business. Since the data stories are meant to aid bank organisations, three individuals who work in the banking industry were interviewed and a discussion has been conducted within the UCI data context. Input from the interviews described that clients who have not paid their credit card payment in seven or eight months will be pursued by banks. Banks also distinguish between excellent and bad-paying masters based on payment trends. This is to track down clients and allow for reminders to be sent out. Furthermore, the bank also would like to anticipate the possibility of a default payment for the following month. One of the user requirements is to view data in a meaningful format, as users will need to identify clients by individual records.

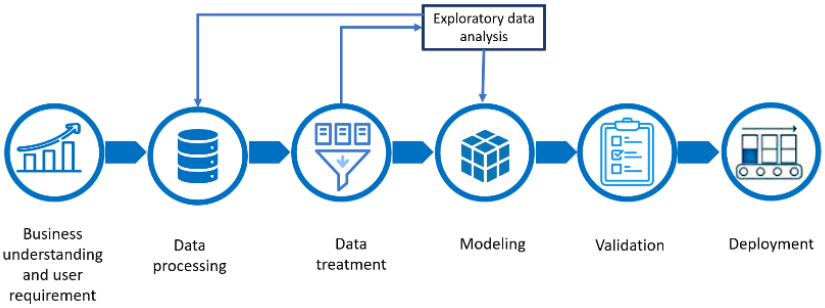


Figure 1. Steps in the study

Table 1. Summary of data dictionary

Variable	Data Type	Level of Measurement
Education	Alphabetic	Ordinal
Marital status and repayment status (September, August, July, June, May, April)	Alphabetic	Nominal
Age, credit limit provided, amount of bill statement (September, August, July, June, May, April)	Numeric	Continuous
Category for gender, default payment next month	Numeric	Binary

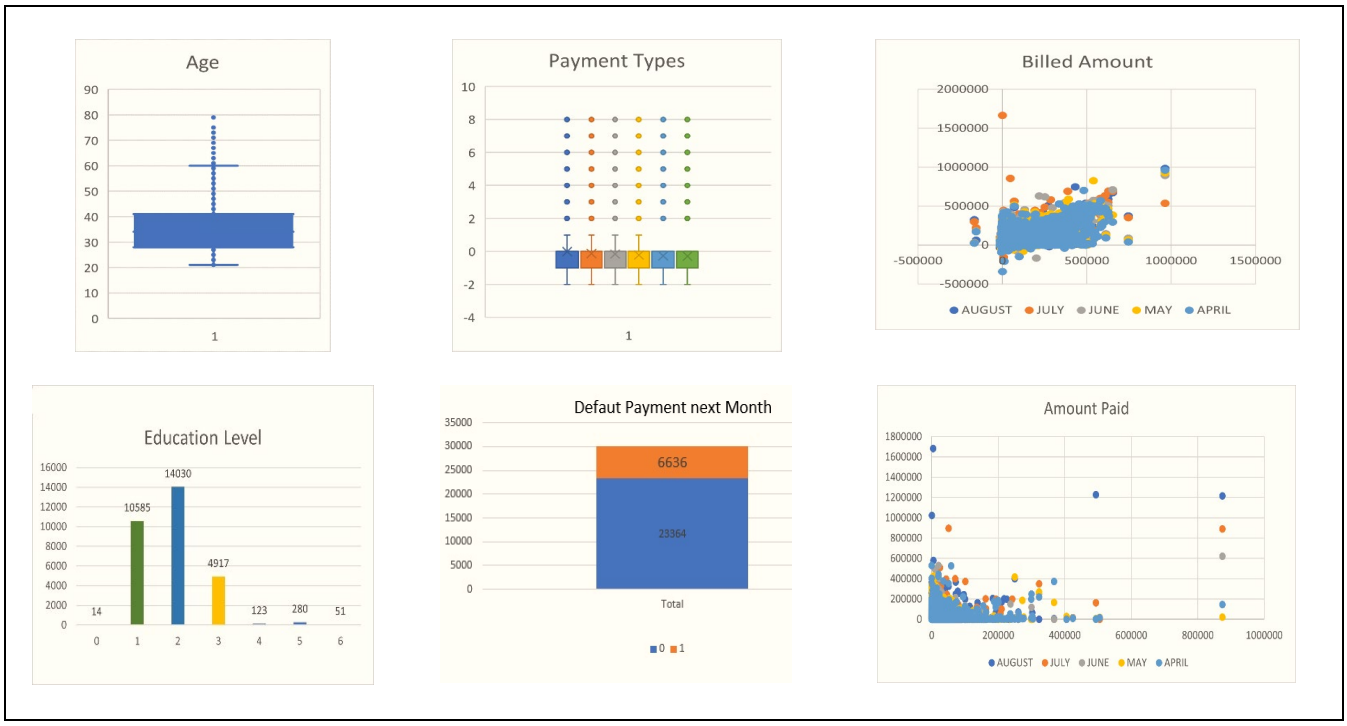
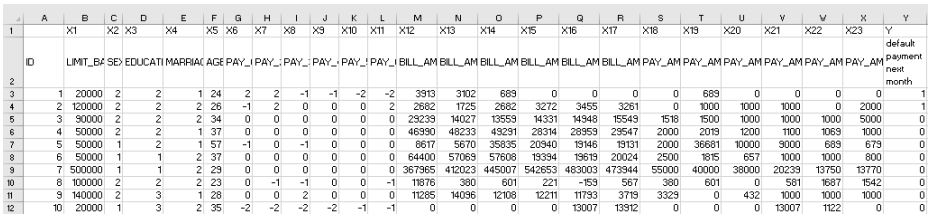


Figure 2. Visual outcome of the EDA

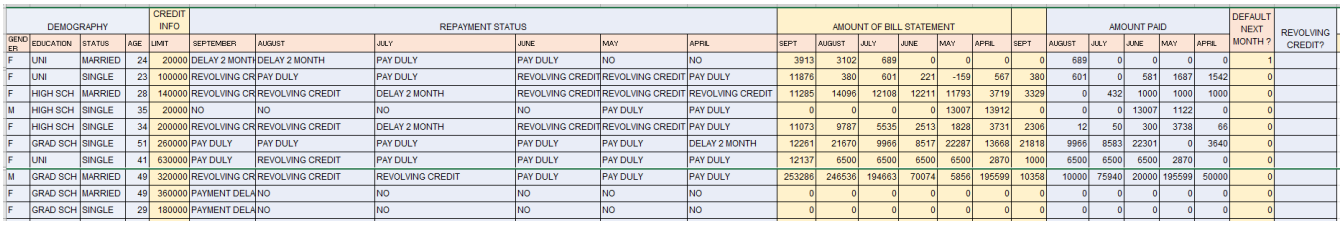
In step 2, the data exploration was conducted. It is found that there were no missing data nor errors in data value. Table 1 shows the summary of the data dictionary while Figure 2 shows the visual of the exploratory data analysis (EDA). The box plot shows that the range of age is between 21 and 79. The education level is between high school, university, graduate school and others. Payment types and codes are as follows: 0 means no consumption, -1 means paying duly, 0 means use of revolving credit, 1 delay means 1 month, and -2 to -8 means payment delay 2-8 months, respectively [26]. The billed amount and the amount paid are shown in scatter charts, where it can be seen that the range and density of the amount paid are related. The stack chart for default payment shows that 6636 clients (value = 1) have defaulted on payment in the following month. In terms of data pre-processing, little effort has been done to clean the data set because it is complete and has been cross-checked against previous work [26]. However, as users (based on the interview) would like to view the data in readable form, data values have been tokenized into meaningful labels. Figure 3 and 4 show the snippets of the dataset.

Step 3 involved data preparation to create visual objects, especially the Sankey chart, which cannot be created if some of the variables are not combined. 12 variables have been chosen to form a Sankey chart, as these variables are required to compose a story that relates the client's credit limit to how many months they have failed to pay and the potential to default in payment in the next month. The profile of the client, like their age group and education level, can also be identified. Table 2 lists the outcome of the pre-processing where 3 derivative variables have been produced. The final step is Step 4, where the visualisation model and dashboard development took place.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1		X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	Y
2	ID	LIMIT_BA	SE	EDUCATI	MARRIAG	AGE	PAY_(	PAY_:	PAY_:	PAY_4	PAY_!	PAY_(	BILL_AM	BILL_AM	BILL_AM	BILL_AM	BILL_AM	BILL_AM	PAY_AM	PAY_AM	PAY_AM	PAY_AM	PAY_AM	PAY_AM	default payment next month
3	1	20000	2	2	1	24	2	2	-1	-1	-2	-2	3913	3102	689	0	0	0	0	689	0	0	0	0	1
4	2	120000	2	2	2	26	-1	2	0	0	0	2	2682	1725	2682	3272	3455	3261	0	1000	1000	1000	0	2000	1
5	3	90000	2	2	2	34	0	0	0	0	0	0	29239	14027	13559	14331	14948	15549	1518	1500	1000	1000	1000	5000	0
6	4	50000	2	2	1	37	0	0	0	0	0	0	46990	48233	49291	28314	28959	29547	2000	2019	1200	1100	1069	1000	0
7	5	50000	1	2	1	57	-1	0	-1	0	0	0	8617	5670	35835	20940	19146	19131	2000	36681	10000	9000	689	679	0
8	6	50000	1	1	2	37	0	0	0	0	0	0	64400	57069	57608	19394	19619	20024	2500	1815	657	1000	1000	800	0
9	7	500000	1	1	2	29	0	0	0	0	0	0	367965	412023	445007	542653	483003	473944	55000	40000	38000	20239	13750	13770	0
10	8	100000	2	2	2	23	0	-1	-1	0	0	-1	11876	380	601	221	-159	567	380	601	0	581	1687	1542	0
11	9	140000	2	3	1	28	0	0	2	0	0	0	11285	14096	12108	12211	11793	3719	3329	0	432	1000	1000	1000	0
12	10	20000	1	3	2	35	-2	-2	-2	-2	-1	-1	0	0	0	0	13007	13912	0	0	0	13007	1122	0	0

Figure 3. Dataset before pre-processing



DEMOGRAPHY				CREDIT INFO	REPAYMENT STATUS						AMOUNT OF BILL STATEMENT							AMOUNT PAID						DEFAULT NEXT MONTH ?	REVOLVING CREDIT?
GENDER	EDUCATION	STATUS	AGE	LIMIT	SEPTEMBER	AUGUST	JULY	JUNE	MAY	APRIL	SEPT	AUGUST	JULY	JUNE	MAY	APRIL	SEPT	AUGUST	JULY	JUNE	MAY	APRIL		
F	UNI	MARRIED	24	20000	DELAY 2 MONTH	DELAY 2 MONTH	PAY DULY	PAY DULY	NO	NO	3913	3102	689	0	0	0	0	689	0	0	0	0	1	
F	UNI	SINGLE	23	100000	REVOLVING CR	PAY DULY	PAY DULY	REVOLVING CREDIT	REVOLVING CREDIT	PAY DULY	11876	380	601	221	-159	567	380	601	0	581	1687	1542	0	
F	HIGH SCH	MARRIED	28	140000	REVOLVING CR	REVOLVING CREDIT	DELAY 2 MONTH	REVOLVING CREDIT	REVOLVING CREDIT	REVOLVING CREDIT	11285	14096	12108	12211	11793	3719	3329	0	432	1000	1000	1000	0	
M	HIGH SCH	SINGLE	35	20000	NO	NO	NO	NO	PAY DULY	PAY DULY	0	0	0	0	13007	13912	0	0	0	13007	1122	0	0	
F	HIGH SCH	SINGLE	34	200000	REVOLVING CR	REVOLVING CREDIT	DELAY 2 MONTH	REVOLVING CREDIT	REVOLVING CREDIT	PAY DULY	11073	9787	5535	2513	1828	3731	2306	12	50	300	3738	66	0	
F	GRAD SCH	SINGLE	51	260000	PAY DULY	PAY DULY	PAY DULY	PAY DULY	PAY DULY	DELAY 2 MONTH	12261	21670	9966	8517	22287	13668	21818	9966	8583	22301	0	3640	0	
F	UNI	SINGLE	41	630000	PAY DULY	REVOLVING CREDIT	PAY DULY	PAY DULY	PAY DULY	PAY DULY	12137	6500	6500	6500	6500	2870	1000	6500	6500	6500	2870	0	0	
M	GRAD SCH	MARRIED	49	320000	REVOLVING CR	REVOLVING CREDIT	REVOLVING CREDIT	PAY DULY	PAY DULY	PAY DULY	253286	246536	194663	70074	5856	195599	10358	10000	75940	20000	195599	50000	0	
F	GRAD SCH	MARRIED	49	360000	PAYMENT DELA	NO	NO	NO	NO	NO	0	0	0	0	0	0	0	0	0	0	0	0	0	
F	GRAD SCH	SINGLE	29	180000	PAYMENT DELA	NO	NO	NO	NO	NO	0	0	0	0	0	0	0	0	0	0	0	0	0	

Figure 4. Dataset after pre-processing

Table 2. Derivative variables

Variable	Description
Age group	Data are classified as follows: 21-30, 31-40, 41-50, 51-60, and over 61.
Credit limit by group	The data are divided into eight categories, each with a 10,0000 gap between them
Month and repayment status	Data occurrences have been counted and merged

4. FINDINGS AND DISCUSSION

The findings will be discussed in two subsections. The first section discusses on the dashboard design, and the second section discusses the results of the usability evaluation. Demographics and the default payment visuals are presented on different pages. The dashboard design is shown in Figure 5. The demographic page comprises seven objects that constitute four charts, a filter, and three cards displaying the overall and number of clients based on gender. Objects on the demographic page are composed in accordance with user requirements. Whereas on the default payment page, there are four charts that correspond to it. This study follows Gestalt principles in arranging objects on each page, as this is to avoid jumpy flow and misinterpretations [28]. The purposes of each chart are described in Table 3.

Figure 6 depicts a six-month pattern of payment delays ranging from one to eight months from April to September. The trend shows that many clients (75.63%) were delayed in payment for only two months before dropping off, while only a few clients (7.8%) delayed payment for three to eight months. The rest (16.6%) did not have problems with payment. This line chart is related to the Sankey chart in Figure 7.

The width of the lines in Figure 7 is significantly related to the number of records they contain. The size of a data cluster is represented by the height of a block, which in this case shows the number of clients. The thickness of the stream represents the number of records contained. Thus, in repayment axes, it can be seen that revolving credit facilities (%) save many of the clients. Its merits lie in its ability to visualise complicated or detailed data and in telling how it moves from one vertical axis to another. The Sankey chart helps to show that a large number of clients are actually using the revolving credit to settle the billed amount. If we look in detail, the understanding of the relationship between the credit limit issued, the months, repayment status, and the chances of defaulting the following month are shown clearly in the flow. The rightmost axis, named "Default", is either 1(default) or 0 (no default), which means the client may or may not have default payment issues in the following month after September. It shows that the majority of the clients (28.4%) will not have default payments in the next month as most of them will be using the revolving credit that will help them. In future work, values for these axes can be obtained from the prediction algorithm. Apart from identifying the numbers of clients of each of the lines, Sankey charts also show individual traces. It can be seen that the largest percentage (71%) of the clients who do not have default payments in the coming month are those who are using revolving credit, paying duly, or those who do not consume credit card. When we trace repayment status by age and marital status, clients between the ages of 21 and 30 years old and single have a greater likelihood of defaulting again for the next month if they delay the repayment for 6 to 8 months. The pattern is when clients who have been late for 6 months or more are more likely to continue delaying for the next month since they already have difficulties paying their outstanding payments. The charts also aid in determining the clients' profiles.

For the dashboard usability assessment, this study employs the Post-Study System Usability Questionnaire [28]. System usefulness, information quality, and interface quality are the three sub-scales. The usefulness of a system is defined as easy to utilise and discover, as well as its ability to execute tasks productively and effectively. The quality of information provided by a system to a user, such as a notification and instructions on how to fix the issue, is referred to as information quality. The level of satisfaction provided by the system to its users is referred to as interface quality. There were 12 respondents involved in the assessment, where 5 were from the banking industry, and 7 were managers from manufacturing, social business, and

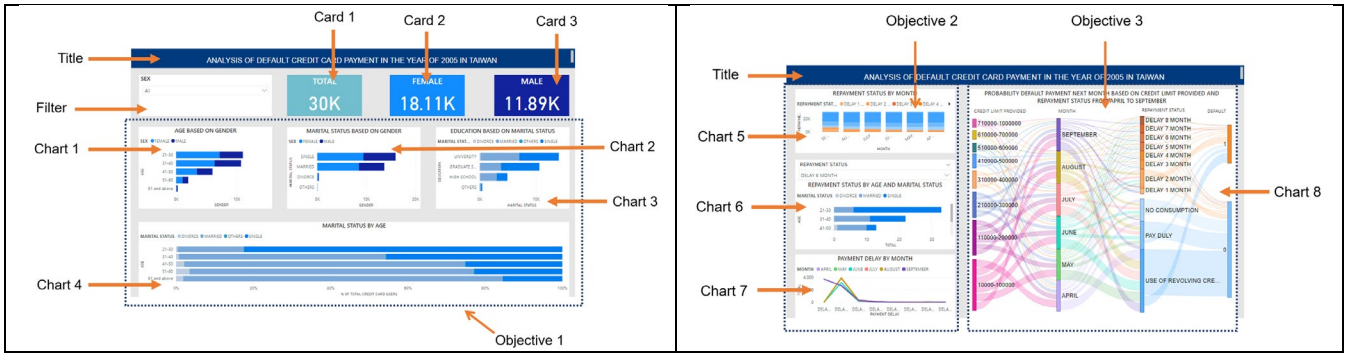


Figure 5. Screenshot of demographic page (left) and default payment page (right)

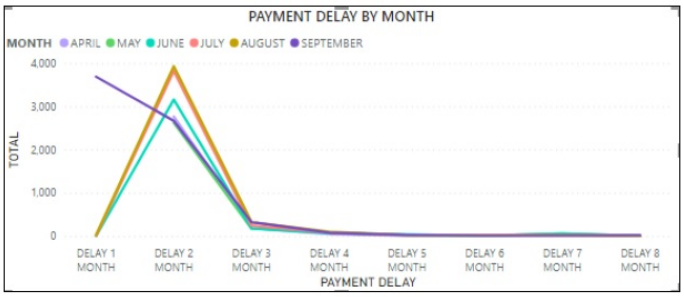


Figure 6. Payment delay by month

Table 3. Description and purpose of the object

Description	Purpose
Client Profiles	
1. Age based on gender 2. Marital status based on gender 3. Education based on marital status 4. Marital status by age	To identify clients' profiles based on age and gender, marital status and gender, client educational level and marital status, and group of age
Repayment Behaviour	
1. Repayment status by month	To determine the frequency and types of late payments that clients experience each month
2. Repayment status by age and marital status	To identify clients with a personal background of 6th to 8th month delays, as they are likely to be delayed again in 9th months
3. Payment delay by month	To identify the trend in payments by month
4. From April to September, the default payment for the next month depends on the credit limit and repayment status.	To see the link and pattern of client actions in relation to the likelihood of default payment in the coming month. Payments in October, for example, may be viewed depending on the credit limit and repayment status from April to September

education sectors who have used dashboards in the workplace. They were given five activities to complete before filling out the questionnaire, all of which were to find and obtain information from the dashboard. The summary of the assessment based on the items is shown in Table 4.

Nine respondents have more than 11 years of working experience. The five-point Likert scale was used, with strongly disagree = 1, disagree = 2, neutral = 3, agree = 4, and strongly agree = 5. Table 4 indicates that the average score ranges from 3.9 to 4.4, signifying that the level of usability is between agree and strongly agree. The 0.7 standard deviation represents the datapoint's variations over a wide range of values. The overall mean is 4.0, indicating that respondents agree that the dashboard is usable in identifying client default payment patterns.

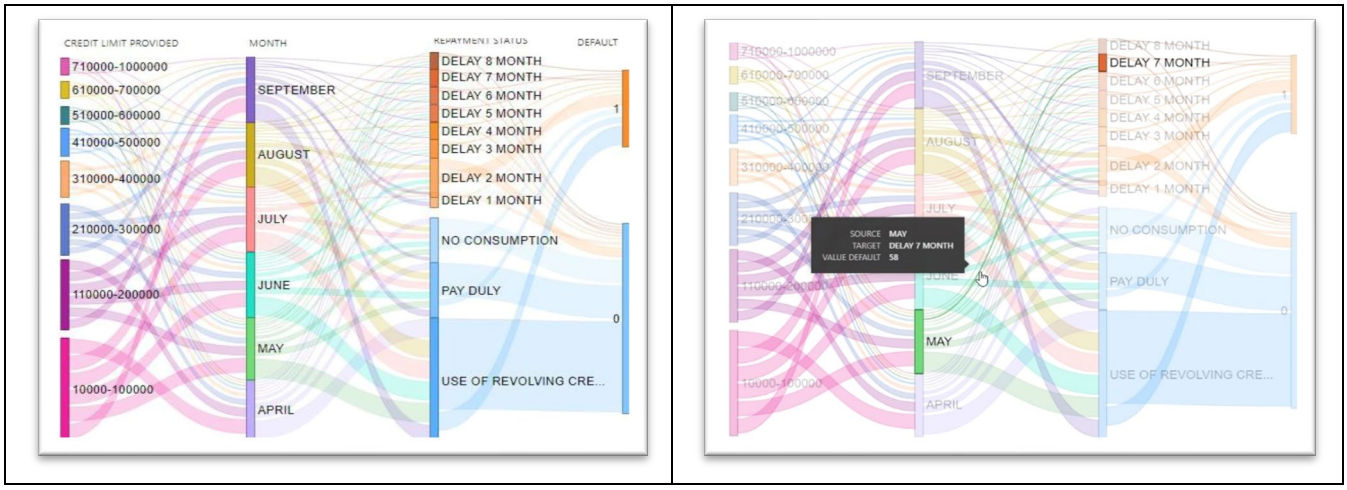


Figure 7. Sankey chart in ideal view (left) and upon clicking (right)

Table 4. Result of the usability test (adapted from [1])

	n = 12			
Item	Mean	Std Dev	Min	Max
The dashboard is necessary and beneficial for identifying credit card default issues	4.0	0.6	3	5
4.0	4.0	0.8	3	5
The dashboard has been clearly designed with the goal of completing the credit card default simple analysis	3.9	0.8	3	5
I am satisfied with the design and features	3.8	0.6	3	5
The data in this dashboard was clear and straightforward	4.4	0.7	3	5
The dashboard assists users (banks/agencies) in understanding pattern and issues	4.3	0.6	3	5
User is able to complete their task by utilising the information contained in this dashboard	4.0	0.6	3	5
It was easy to find the data/information required by the user	4.1	0.8	2	5
The interface was pleasant to use	3.8	0.8	3	5
The dashboard enables the users to analyse data	3.9	0.8	3	5
It includes all of the features and capabilities necessary	3.7	0.7	3	5
In general, I'm pleased with the dashboard	4.0	0.9	3	5
There are a few visual elements that could be improved	4.3	0.5	4	5
Average Mean	4.0	0.7	-	-

5. CONCLUSION

This study simulated the development of a dashboard that assists users at credit agencies or banking institutions in detecting credit card payment default. In this research, Sankey chart was used to examine patterns based on delayed patterns, as well as the usage of revolving credit, which helps most clients pay their bills. The dashboard is usable according to the assessment, however there are a few visual elements that could be enhanced. In future work, a prediction model will be constructed and integrated into the default payment indicator.

REFERENCES

[1] A. S. Khalid, N. H. Hassan, N. A. A. B. Razak and A. F. Baharuden, Business intelligence dashboard for driver performance, in fleet management, *International Conference on E-Education, E-Business, E-Management, and E-Learning,* Osaka, Japan, 2020, 347-351.

[2] S. Few, *Information Dashboard Design: Effective Visual Communication of Data*, O'Really Media, 2006.

[3] J. Cai, X. Liu and Y. Wu, SVM learning for default prediction of credit card under differential privacy, *Workshop on Privacy-Preserving Machine Learning in Practice*, 2020, 3 pages.

[4] C. J. Costa and M. Aparicio, Supporting the decision on dashboard design Charts, *Proceeding of 254th IIER International Conference,* Saint Petersburg, Russia, 2019, 10-15.

[5] W. W. Eckerson, *Performance Dashboards. Measuring, Monitoring, and Managing Your Business*, Hoboken: Wiley, 2011.

[6] R. K. Rainer and C. G. Cegielski, *Introduction to Information Systems. Supporting and Transforming Business*, Hoboken: Wiley, 2011.

[7] A. Janes, A. Sillitti and G. Succi, Effective dashboard design, *Cutter IT Journal*, 26(1), 2013, 17-24.

[8] S. Sharma and V. Mehra, *Default payment analysis of credit card clients*, 2018, 7 pages, DOI: 10.13140/RG.2.2.31307.28967.

[9] M. M. Hassan and T. Mirza, Credit card default prediction using artificial neural networks, *GIS Science Journal*, 7, 2020, 383-390.

[10] Y. Ma, Prediction of default probability of credit-card bills, *Open Journal of Business and Management*, 8, 2020, 231-244.
[11] R. L. Liu, Machine learning approaches to predict default of credit card clients. *Modern Economy*, 9, 2018, 1828-1838.
[12] J. Zhou, W. Li, J. Wang, S. Ding and C. Xia, Default prediction in P2P lending from high-dimensional data based on machine learning, *Physica A: Statistical Mechanics and its Applications,* 534, 2019, 122370.
[13] Y. Xia, L. He, Y. Li, N. Liu and Y. Ding, Predicting loan default in peer-to-peer lending using narrative data, *Journal of Forecasting*, 39(2), 2020, 260-280.
[14] T. M. Alam, K. Shaukat, I. A. Hameed, S. Luo, M. U. Sarwar, S. Shabbir, J. Li and M. Khushi, An investigation of credit card default prediction in the imbalanced datasets, *IEEE Access*, 8, 2020, 201173-201197.
[15] O. J. Leong and M. Jayabalan, A comparative study on credit card default risk predictive model, *Journal of Computer Theoretical Nanoscience*, 16(8), 2019, 3591-3595.
[16] A. Hasbun, *An Empirical Investigation: Do Animated Graphs Improve the Quality of Sales Forecasting Decisions in Comparison to Tables?* Master's Thesis, Hanken School of Economics, 2009.
[17] K. D. Lawrence, S. Kudyba and R. K. Klimberg, *Data Mining Methods and Applications*, CRC Press, 2008.
[18] W. Eckerson and M. Hammond, *Visual Reporting and Analysis: Seeing Is Knowing*, TDWI-The Data Warehousing Institute, 2011, 1-24.
[19] L. Pappas and L. Whitman, Riding the technology wave: effective dashboard data visualization, in *Human Interface and the Management of Information,* M. J. Smith and G. Salvendy (eds), Lecture Notes in Computer Science, 6771011, 249-258.
[20] A. Chaudhuri, *A Visual Technique to Analyze Flow of Information in a Machine Learning System*, Walmart Labs, USA, 2019.
[21] A. Lamer, G. Laurent, S. Pelayo, M. E. Amrani, E. Chazard and R. Marcilly, Exploring patient path through sankey diagram: A proof of concept, *Studies in Health Technology and Informatics*, 270, 2020, 218-222.
[22] V. R. Basili, M. Lindvall, M. Regardie, C. Seaman, J. Heidrich, J. Munch, D. Rombach and A. Trendowicz, Linking software development and business strategy through measurement, *Computer*, 43(4), 2010, 57-65.
[23] S. Mazumdar, A. Varga, V. Lanfranchi, D. Petrelli and F. Ciravegna, A knowledge dashboard for manufacturing industries, *The Semantic Web: ESWC 2011 Workshops*, Lecture Notes in Computer Science, 7117, 2012, 112-124.
[24] J. Li, J. Liu, W. Xu and Y. Shi, Support vector machines approach to credit assessment, *International Conference on Computational Science*, Lecture Notes in Computer Science, 3039, 2004, 892-899.
[25] K. Bache and M. Lichman, *UCI Machine Learning Repository*, University of California, School of Information and Computer Science, 2013.
[26] I. -C. Yeh and C. -H. Lien, The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients, *Expert Systems with Applications*, 36(2), 2009, 2473-2480.
[27] S. Guberman, On Gestalt theory principles, *Gestalt Theory*, 37(1), 2015, 25-44.
[28] A. L. Fruhling and S. Lee, Assessing the reliability, validity and adaptability of PSSUQ, *11th Americas Conference on Information Systems*, Nebraska, USA, 2005, 2394-2402.