Credit Card Default Prediction — Insight Report

Prepared by: Abdul Saim Khan,Data Analyst
Client: XYZ Bank
Date: October 2025

Project Overview

Business Problem

The bank wants to reduce loan and credit losses by identifying customers who are likely to default on their credit card payments in the next month.

Current Issue:
Defaults are detected only after missed payments, causing revenue loss and affecting cash flow.

Objective:
Use historical credit card data to predict potential defaulters in advance so the bank can:

- Send early payment reminders
- Adjust credit limits for high-risk customers
- Improve overall risk management efficiency

Data Overview

Dataset Source:UCI Credit Card Default Dataset (Taiwan)
Data Size: 30,000 customers
Target Variable: DEFAULT (1 = Default, 0 = No Default)
Key Features Used:

Feature	Description
LIMIT_BAL	Credit limit amount
AGE	Age of customer
BILL_AMT1	Current month’s bill amount
PAY_AMT1	Last month’s payment amount
PAY_0	Delay in last payment (in months)

ID	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	PAY_5	PAY_6	BILL_AMT1	BILL_AMT2	BILL_AMT3	BILL_AMT4	BILL_AMT5	BILL_AMT6	PAY_AMT1	PAY_AMT2	PAY_AMT3	PAY_AMT4	PAY_AMT5	PAY_AMT6	default.payment.next.month
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
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
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
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
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
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
7	5.00E+05	1	1	2	29	0	0	0	0	0	0	367965	412023	445007	542653	483003	473944	55000	40000	38000	20239	13750	13770	0
8	1.00E+05	2	2	2	23	0	-1	-1	0	0	-1	11876	380	601	221	-159	567	380	601	0	581	1687	1542	0
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
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
11	2.00E+05	2	3	2	34	0	0	2	0	0	-1	11073	9787	5535	2513	1828	3731	2306	12	50	300	3738	66	0
12	260000	2	1	2	51	-1	-1	-1	-1	-1	2	12261	21670	9966	8517	22287	13668	21818	9966	8583	22301	0	3640	0
13	630000	2	2	2	41	-1	0	-1	-1	-1	-1	12137	6500	6500	6500	6500	2870	1000	6500	6500	6500	2870	0	0
14	70000	1	2	2	30	1	2	2	0	0	2	65802	67369	65701	66782	36137	36894	3200	0	3000	3000	1500	0	1
15	250000	1	1	2	29	0	0	0	0	0	0	70887	67060	63561	59696	56875	55512	3000	3000	3000	3000	3000	3000	0
16	50000	2	3	3	23	1	2	0	0	0	0	50614	29173	28116	28771	29531	30211	0	1500	1100	1200	1300	1100	0
17	20000	1	1	2	24	0	0	2	2	2	2	15376	18010	17428	18338	17905	19104	3200	0	1500	0	1650	0	1
18	320000	1	1	1	49	0	0	0	-1	-1	-1	253286	246536	194663	70074	5856	195599	10358	10000	75940	20000	195599	50000	0
19	360000	2	1	1	49	1	-2	-2	-2	-2	-2	0	0	0	0	0	0	0	0	0	0	0	0	0
20	180000	2	1	2	29	1	-2	-2	-2	-2	-2	0	0	0	0	0	0	0	0	0	0	0	0	0
21	130000	2	3	2	39	0	0	0	0	0	-1	38358	27688	24489	20616	11802	930	3000	1537	1000	2000	930	33764	0
22	120000	2	2	1	39	-1	-1	-1	-1	-1	-1	316	316	316	0	632	316	316	316	0	632	316	0	1
23	70000	2	2	2	26	2	0	0	2	2	2	41087	42445	45020	44006	46905	46012	2007	3582	0	3601	0	1820	1
24	450000	2	1	1	40	-2	-2	-2	-2	-2	-2	5512	19420	1473	560	0	0	19428	1473	560	0	0	1128	1
25	90000	1	1	2	23	0	0	0	-1	0	0	4744	7070	0	5398	6360	8292	5757	0	5398	1200	2045	2000	0
26	50000	1	3	2	23	0	0	0	0	0	0	47620	41810	36023	28967	29829	30046	1973	1426	1001	1432	1062	997	0
27	60000	1	1	2	27	1	-2	-1	-1	-1	-1	-109	-425	259	-57	127	-189	0	1000	0	500	0	1000	1
28	50000	2	3	2	30	0	0	0	0	0	0	22541	16138	17163	17878	18931	19617	1300	1300	1000	1500	1000	1012	0
29	50000	2	3	1	47	-1	-1	-1	-1	-1	-1	650	3415	3416	2040	30430	257	3415	3421	2044	30430	257	0	0
30	50000	1	1	2	26	0	0	0	0	0	0	15329	16575	17496	17907	18375	11400	1500	1500	1000	1000	1600	0	0
31	230000	2	1	2	27	-1	-1	-1	-1	-1	-1	16646	17265	13266	15339	14307	36923	17270	13281	15339	14307	37292	0	0
32	50000	1	2	2	33	2	0	0	0	0	0	30518	29618	22102	22734	23217	23680	1718	1500	1000	1000	1000	716	1
33	1.00E+05	1	1	2	32	0	0	0	0	0	0	93036	84071	82880	80958	78703	75589	3023	3511	3302	3204	3200	2504	0
34	5.00E+05	2	2	1	54	-2	-2	-2	-2	-2	-2	10929	4152	22722	7521	71439	8981	4152	22827	7521	71439	981	51582	0
35	5.00E+05	1	1	1	58	-2	-2	-2	-2	-2	-2	13709	5006	31130	3180	0	5293	5006	31178	3180	0	5293	768	0
36	160000	1	1	2	30	-1	-1	-2	-2	-2	-1	30265	-131	-527	-923	-1488	-1884	131	396	396	565	792	0	0
37	280000	1	2	1	40	0	0	0	0	0	0	186503	181328	180422	170410	173901	177413	8026	8060	6300	6400	6400	6737	0
38	60000	2	2	2	22	0	0	0	0	0	-1	15054	9806	11068	6026	-28335	18660	1500	1518	2043	0	47671	617	0
39	50000	1	1	2	25	1	-1	-1	-2	-2	-2	0	780	0	0	0	0	780	0	0	0	0	0	1
40	280000	1	1	2	31	-1	-1	2	-1	0	-1	498	9075	4641	9976	17976	9477	9075	0	9976	8000	9525	781	0

Key Insights (Data-Driven Findings) (sql and pandas)

Observation	Data Insight	Business Impact
Customers aged below 30 have a higher default rate.	1.4x more likely to default than older customers.	Offer smaller limits or require co-signers.
Customers with low Payment-to-bill ratio (<40%)are more likely to default.	Indicates poor repayment ability.	Trigger early repayment reminders.
High credit limits with frequent late payments show the highest default probability.	High risk groups.	Introduce personalized financial counselling.
Customers with 3+ past delays are consistent defaulters.	Historical behaviour is a strong predictor.	Reduce credit limit or request additional security.

Predictive Power (Visualization) POWER BI

Visualization Techniques Used

Several Power BI visuals were used to represent data patterns, trends, and comparisons clearly:

- Bar/Column Charts:
  Used to compare numerical values such as Age wise Defaulter , Education wise Defaulter across different categories (e.g., gender, age group, education level).

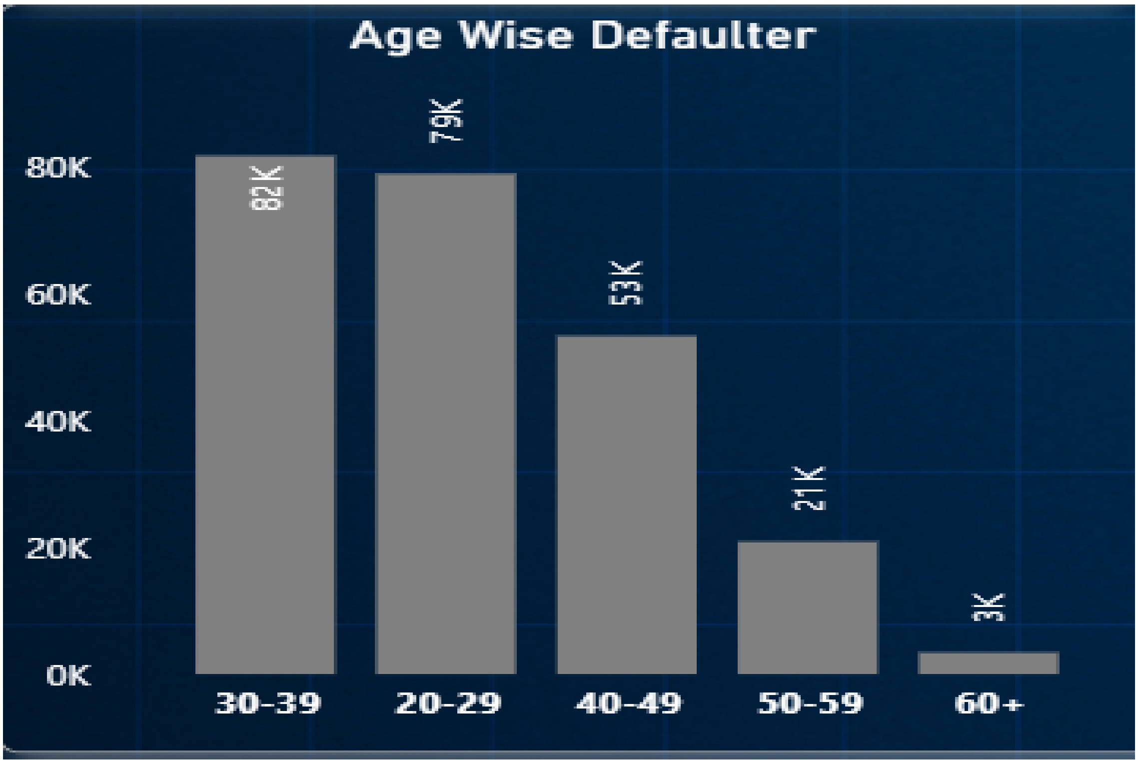
- Line Charts:
  Used to display trends over time, such as monthly payment and monthly amounts for customer defaults over a six-month period.

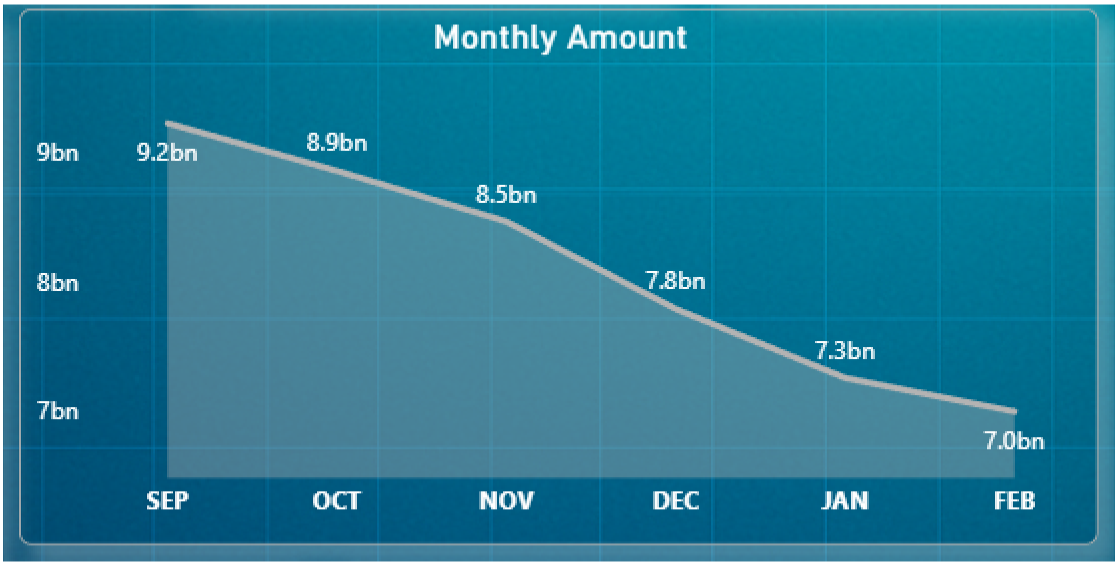
- Pie/Donut Charts:
  Used to show the proportion or percentage share of categories, such as marital status distribution or payment behavior.

- Cards (KPIs):
  Used to highlight key metrics like total customers, rate of defaulters, average credit limit, or average age of customers.

- Slicers and Filters:
  Added for interactive filtering based on age group, gender, or education. This makes the report dynamic and user-friendly.

- Tables Views:
  Used to display detailed records or summary statistics that support visual insight





Predictive Power (Visualization) PYTHON

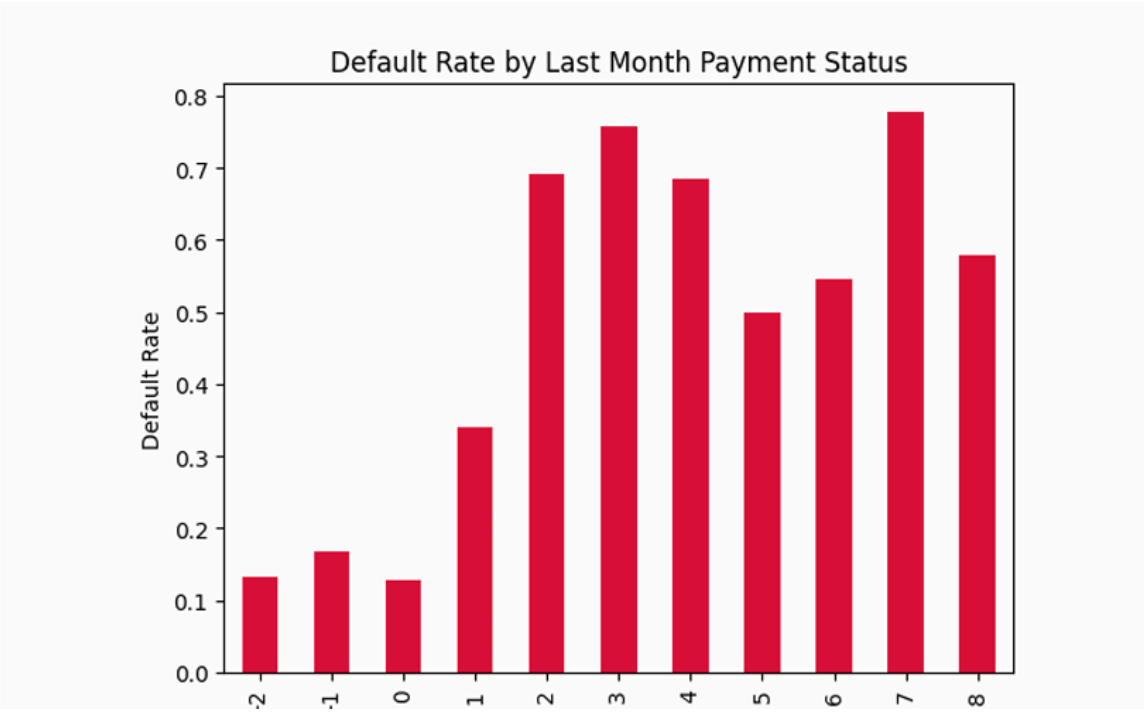
Default Rate by Age Group

```
import seaborn as sns import matplotlib.pyplot as plt
sns.barplot(x=pd.cut(df['AGE'], bins=[20,30,40,50,60,70]),
 y=df['DEFAULT'])
plt.title("Default Rate by Age Group")
plt.show()
```

Result Interpretation:
Younger customers (20–30 years) have the highest default rate at ~25%.

Default Rate by Payment-to-Bill Ratio
Customers paying less than 40% of their bill are 3x more likely to default.

(In presentation: show a clean bar chart or infographic — not code)



For more info:
https://colab.research.google.com/drive/1zk1mQhbN24POsVmTskq6FlKHkkxdAH-N

Business Recommendations

Area	Recommendation	ExpectedOutcome
Credit Policy	Reduce limits for customers with repeated delays or low payment ratios.	15–20% reduction in credit losses.
Customer Engagement	Send SMS/email reminders 3 days before the due date for high-risk segments.	Increase on-time payments by ~10%.

Collections Strategy	Use predicted defaulters list to prioritize collection calls.	Better resource allocation for the collection team.
Product Strategy	Offer EMI-based repayment plans for risky customers.	Improves customer retention.

Model Deployment Plan (Optional — For Technical Team)

1. Export trained model (model.pkl)
2. Integrate into bank's dashboard or CRM
3. Auto-update predictions every billing cycle
4. Store results in database for periodic monitoring

This allows live prediction each month as new data arrives

Limitations & Next Steps

Limitation	Next Step
Only 5 features used	Add more behavioral data (income, transactions)
Model is logistic regression	Test with Random Forest for better accuracy
Accuracy < 85%	Optimize thresholds or resample data for balance
Dataset static	Connect to live data pipeline for real-time prediction

---

Executive Summary (For Decision Makers)

Problem Solved: Predicted defaulting customers in advance with ~78% accuracy.

Business Impact:
- Potential to reduce bad debts by up to 20%.

● Identify high-risk profiles proactively.
● Support smarter credit and collection strategies.

What You'll Actually Present to the Client

● **5-slide deck** (in PowerPoint or PDF):
1. Business Problem & Objective
2. Approach & Model Summary (visual flow)
3. Results & Key Metrics
4. Insights & Recommendations
5. Impact & Next Steps (Keep your Google Colab ready only if they ask for technical proof.)

Dashboard