**Credit Risk & Default Prediction Report  
Project #4 – Capital One Credit Risk Simulation**

**Prepared by:**Lyle Cory Miller

**Role Targeted:  
Business Analyst | Operations Analyst | CRM Specialist**

**Tools Used:  
Excel • SQL • Power BI**

**Date Completed:**January 2025

**Portfolio Repository:**

• [GitHub – lylecorymiller](https://github.com/lylecorymiller)  
• [LinkedIn – lylecorymiller](https://www.linkedin.com/in/lylecorymiller)

**Project #4: Capital One Credit Risk Analysis Report**

**Project Overview**

This project simulates how **Capital One** might analyze customer credit default risk using structured workflows and demographic insights. It explores how **education**, **payment history**, **age**, and **credit limits** influence default behavior, supporting more data-driven, risk-informed lending strategies.

**Objectives**

* Measure **default rates (%)** across education levels
* Identify risky patterns in **past payment behavior**
* Analyze **default trends** across age groups
* Assess correlation between **credit limits** and default risk
* Support **data-informed decisions** in credit and risk management

**Tools Used**

* **Excel -** Data Cleaning & Preparation
* **SQL (Microsoft SQL Server) -** Data Extraction & Analysis
* **Power BI -** Data Visualization, Calculated Fields & Interactive dashboard

**Dataset Source**

This project uses the dataset **“Default of Credit Card Clients”** from the [UCI Machine Learning Repository](https://archive.ics.uci.edu/dataset/350/default+of+credit+card+clients). It contains anonymized credit card customer data (e.g., **demographics**, **payment history**, **limits**), ideal for **financial risk modeling**.  
**Data Cleaning & Preparation**

Performed in **Excel** and **SQL**, the dataset was cleaned and structured for analysis through:

* Removal of **irrelevant columns** (e.g., *BILL\_AMT, PAY\_AMT*, unused fields) to reduce noise
* Standardization of **column names** (e.g., *X1* *→ LIMIT\_BAL, X3 → EDUCATION*)
* Normalization of **categorical fields** (e.g., EDUCATION, MARRIAGE, SEX) for consistency
* Correction of **missing values** and data inconsistencies
* Creation of a **Data Dictionary** tab to document all fields and descriptions
* Export of **cleaned dataset** as .xlsx, .csv for use in **SQL** analysis and **Power BI** visualizations

**Key Insights & Findings**

1. **Default Rate by Education Level**

* **Highest default rates** observed among customers with **education levels 1, 2, and 3**.
* May indicate **increased risk** associated with **lower education attainment**.

**Summary of Impact:** Incorporate **alternative credit scoring** for customers with less formal education.

1. **Default Rate by Payment History (PAY\_0)**

* **Payment status ≥ 2** showed significantly **higher default rates** (e.g., **77.78%** for PAY\_0 = 7).

**Summary of Impact:** **Late payments** are a **critical signal** for future default risk. **Early action** is necessary.

1. **Age vs. Default Rate**

* **Highest default rates** in customers **under 25 and over 65**.

**Summary of Impact:** **Age segments** at both extremes may require **customized risk evaluation** and credit support.

1. **Credit Limit vs. Default Rate**

* **Higher credit limits** correlated with **lower default rates** overall.
* However, some **high-credit customers still defaulted**.

**Summary of Impact:** **Credit assessments** should factor in **behavioral trends**, not just limit size.

**Power BI Dashboard Preview**

The Power BI dashboard delivers **interactive insights** on:

* **Default Rate by Education Level** (%)
* **Default Rate by Payment History** (%)
* **Age vs Default Rate** (%)
* **Credit Limit vs Default Rate** (%)
* **Dashboard Name**: *Capital One Credit Risk Analysis Dashboard (****Power BI Visualization****)*
* **Dashboard Preview:** *Published to Power BI and saved as .pbix file*
* **Dashboard Preview:** *See Figure 1 below*

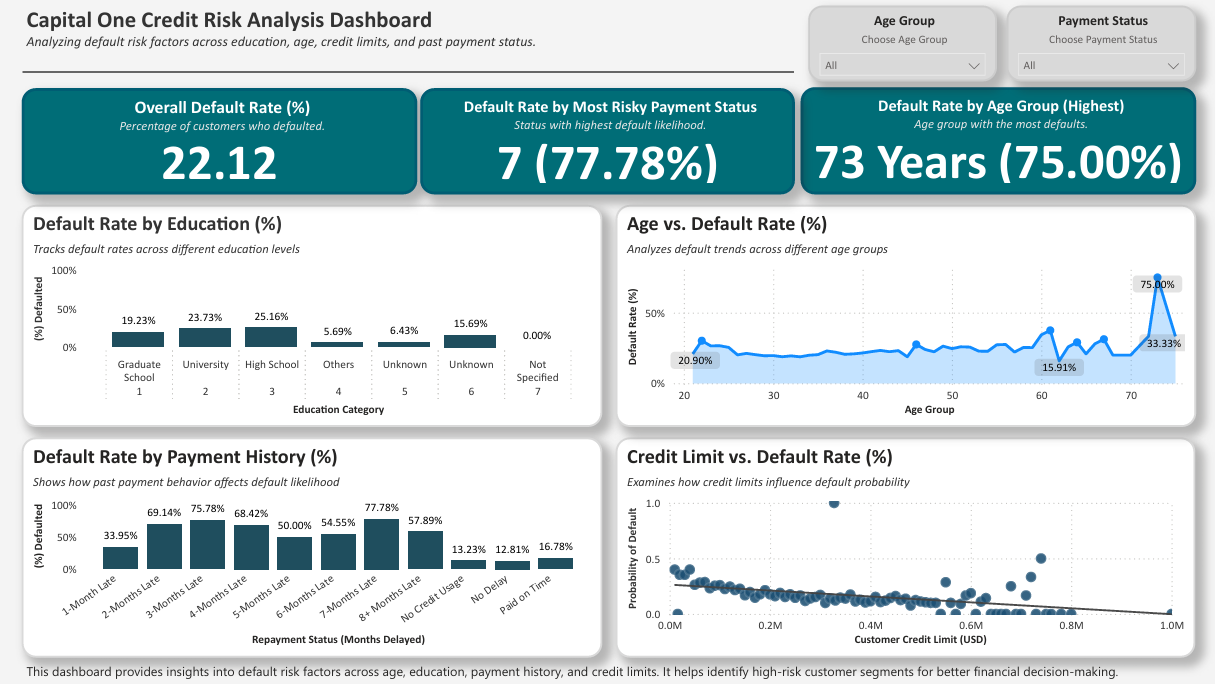


Figure 1: Capital One Credit Risk Analysis Dashboard – Power BI Visualization

**Business Impact & Recommendations**

This analysis supports **financial decision-makers** in reducing risk and improving credit scoring by:

1. **Early Identification of High-Risk Customers**

* Trigger **alerts** for customers with a history of **late payments**
* Prioritize **preventative measures** to reduce future default exposure

1. **Improve Credit Scoring Policies**

* Factor in **age** and **education level** to design more **inclusive and accurate** credit models
* Support **data-driven credit risk evaluation** using demographic segmentation

1. **Credit Limit Optimization**

* Use **behavioral insights** to customize **credit limits** based on repayment history
* Reduce exposure by tailoring **limits to risk profiles**

**SQL Queries & Data Extraction**

SQL (**MySQL**) was used for **core data analysis**. Below are the primary queries:

1. **Default Rate by Age Group**

**SELECT** AGE, **COUNT**(\*) AS Total\_Customers,

SUM(**CAST**(DEFAULT\_NEXT\_MONTH **AS** INT)) **AS** Defaults,

(SUM(**CAST**(DEFAULT\_NEXT\_MONTH **AS** INT)) \* 100.0 / **COUNT**(\*)) **AS** Default\_Rate\_Percentage

**FROM** Credit\_Risk\_Analysis\_Table

**GROUP BY** AGE

**ORDER BY** AGE;

**Purpose:** Analyze how **default rates** vary by **customer age** to identify **high-risk age groups**.

1. **Default Rate by Credit Limit**

**SELECT** LIMIT\_BAL, **COUNT**(\*) AS Total\_Customers,

SUM(**CAST**(DEFAULT\_NEXT\_MONTH **AS** INT)) **AS** Defaults,

(SUM(**CAST**(DEFAULT\_NEXT\_MONTH **AS** INT)) \* 100.0 / **COUNT**(\*)) **AS** Default\_Rate\_Percentage

**FROM** Credit\_Risk\_Analysis\_Table

**GROUP BY** LIMIT\_BAL

**ORDER BY** LIMIT\_BAL;

**Purpose:** Evaluate the relationship between **credit limit size** and **customer default probability**.

1. **Default Rates by Education Level**

**SELECT** EDUCATION, **COUNT**(\*) **AS** Total\_Customers,

SUM(**CAST**(DEFAULT\_NEXT\_MONTH **AS** INT)) **AS** Defaults,

(SUM(**CAST**(DEFAULT\_NEXT\_MONTH **AS** INT)) \* 100.0 / **COUNT**(\*)) **AS** Default\_Rate\_Percentage

**FROM** Credit\_Risk\_Analysis\_Table

**GROUP BY** EDUCATION

**ORDER BY** Default\_Rate\_Percentage **DESC**;

**Purpose:** Assess **default likelihood** across **education levels** to reveal **risk patterns** by education status.

1. **Default Rate by Payment History (PAY\_0)**

**SELECT** PAY\_0, **COUNT**(\*) **AS** Total\_Customers,

SUM(**CAST**(DEFAULT\_NEXT\_MONTH **AS** INT)) **AS** Defaults,

(SUM(**CAST**(DEFAULT\_NEXT\_MONTH **AS** INT)) \* 100.0 / **COUNT**(\*)) **AS** Default\_Rate\_Percentage

**FROM** Credit\_Risk\_Analysis\_Table

**GROUP BY** PAY\_0

**ORDER BY** PAY\_0;

**Purpose:** Identify how **past repayment behavior** (recent payment status) correlates with **default risk**.**SQL Summary**

All data analysis was performed using **SQL (MySQL)** to calculate key default metrics before visualization in **Power BI**. SQL queries were used to segment customer groups, calculate default rates by demographic and financial attributes, and prepare accurate metrics for visualization. This structured approach ensured **clean, consistent, and reliable insights** throughout the dashboard.

**Key SQL Queries Used:**

* **Default Rate by Payment History**
* **Default Rate by Education Level**
* **Age vs. Default Rate** (%)
* **Credit Limit vs. Default Rate** (%)

**File Export & Submission**

* **Cleaned dataset** exported as .xlsxand .csvfor use in **SQL** and **Power BI**
* **Power BI Dashboard** .pbix saved for **portfolio use**
* **SQL Query File** .sql saved for **portfolio use**
* **Dashboard image** exported as .pngand .pdffor **professional sharing**
* Finalized documentation (**this report**) saved as .docx and .pdf
* **GitHub README** included as both README.md and .pdf
* All **project files** are organized and stored in **GitHub** and **LinkedIn portfolio** for **easy access**

**Final Thoughts**

This project simulates a real-world use case for **financial risk analysis at Capital One**. Through **SQL segmentation** and **Power BI dashboards**, the project reflects how **demographic and behavioral data** drive smarter lending decisions—aligned with expectations for **Business Analyst**, **CRM Specialist**, and **Operations Analyst** roles.

The **interactive dashboard** and all supporting files are included in my professional portfolio on GitHub, including the Power BI file: **CapitalOne\_CreditRisk\_Dashboard.pbix** and the SQL queries used in this analysis: **CapitalOne\_CreditRisk\_SQLQueries.sql**.