📌 ***1. Project Title***

**“Customers’ behavior on credit cards”.**

🎯 ***2. Purpose & Questions***

🔹 Purpose:

* *Identify customer groups that bring the highest value from credit card usage in order to persude the retention strategies and optimizing profits.*

🔹 Key Analytical Questions:

* *What do the customer purchase behavior/value accumulation patterns have in common?*
* *Which customer groups should be prioritized for marketing and retention?*
* *Is there a relationship between demographic characteristics and spending behavior?*

📊 ***3. Data Source & Overview***

🔹 *Data source*: Kaggle from the link [💳 Financial Transactions Dataset: Analytics](https://www.kaggle.com/datasets/computingvictor/transactions-fraud-datasets/data)

🔹 *Data structure*:

* *transactions\_data.csv* – 12 colums and 1,048,575 rows which contains the information of date, card\_id, amount, merchant, mcc, errors, etc.



* *user\_data.csv* – 13 colums and 6,146 rows which has the user data includes gender, age, income, credit\_score, location, etc.



* *card\_data.csv* – 14 colums and 2,000 rows which includes the card details like card\_brand, card\_type, credit\_limit, has\_chip, etc.



* *mcc\_codes.json* – 02 colunms and 109 rows which has the MCC codes and product/service descriptions (Merchant Category Codes)

A screenshot of a menu

AI-generated content may be incorrect.

🧹 ***4. Methodology***

🔹 Data Cleaning

* Process the amount columns containing $ characters, negative data → convert and filter.
* Check and normalize data types (int, datetime, str, etc.)
* Remove records missing important information (NaN, duplicate, transaction error)

🔹 Exploratory Data Analysis (EDA):

* Analyze spending distribution by age group, gender, income.
* Analyze behavior by mcc (industry), card\_type, credit\_limit.
* Statistics on transaction frequency, sales by week/month/year.

🔹 Analysis & Modeling

* Cluster customers by K-Means / Hierarchical Clustering.
* Analyze correlation between credit\_score, income, debt, spending.

🔹 Visualization & Storytelling

* Interactive Dashboard using Power BI
  + Customer distribution map by region.
  + Bar chart analyzing behavior by MCC.
  + Timeline of spending by segment.

📈 ***5. Expected Outcome***

🔹 Key Findings:

* High income customers and good credit scores contributes more than remaining groups.
* Users with card usage for over 5 years are more likely to be loyal.
* MCC related to Dining (5812) and Shopping (5411) account for the majority of repeat transactions.

🔹 Recommendations

* Focus on customers aged 36– 50 with a consistent spending history.
* Implement cashback/shopping offers for popular MCC groups.
* Warn people with low credit scores + spending over the limit.

🧰 ***6. Technologies***

| **Tools** | **Intended us** |
| --- | --- |
| **Python** (pandas, seaborn, scikit-learn) | Data cleaning, EDA, clustering/LTV models |
| **Power BI** / **Plotly Dash** | Plotly Dash Data visualization, presentation dashboards |

📦 ***7. Deliverables***

Code + documentation: sharing to GitHub (README.md)

Final report: PDF/Word to summarize the process and findings.

Slides presentation: storytelling + visual + insight

🕓 ***8. Timeline***

* 26/06/2025 : Selecting the topic + getting data
* 03/07/2025 : Project proposal
* 08-15/07/2025 : Reviewing and Q&A sessions to discuss your progress and receive feedback from lecturers/instructors.
* 15/08/2025 : Uploading the project to GitHub *(after the extension)*
* 18-22/08/2025 : Presenting project to the class and lecturers/instructors.