End-to-End Credit Risk Analysis and Policy Scenario Evaluation

---

Author: Soroush Azarinejadianlangroudi
Role: Junior Data Analyst

1. Executive Summary

This report presents an end-to-end credit risk analysis based on a large-scale consumer lending dataset. The objective is to understand portfolio-level risk, identify high-risk customer segments, and evaluate how changes in credit policy thresholds impact approval volumes and default risk. The analysis combines SQL-based data preparation, analytical modelling in Power BI using DAX, and interactive scenario analysis.

Overall, the portfolio shows a relatively low default rate at an aggregate level. However, risk is not evenly distributed. Defaults are concentrated within specific customer segments, particularly younger customers in lower income bands and customers with a history of previous credit refusals. Scenario analysis further demonstrates clear trade-offs between approval volume and risk when adjusting credit-to-income (CTI) thresholds.

The findings highlight strong opportunities for targeted, risk-based policy optimisation rather than uniform credit tightening.

---

2. Data Overview and Preparation

The analysis is based on a cleaned and model-ready dataset derived from multiple relational tables. The final analytical table was constructed using SQL joins and aggregations, and includes the following key features:

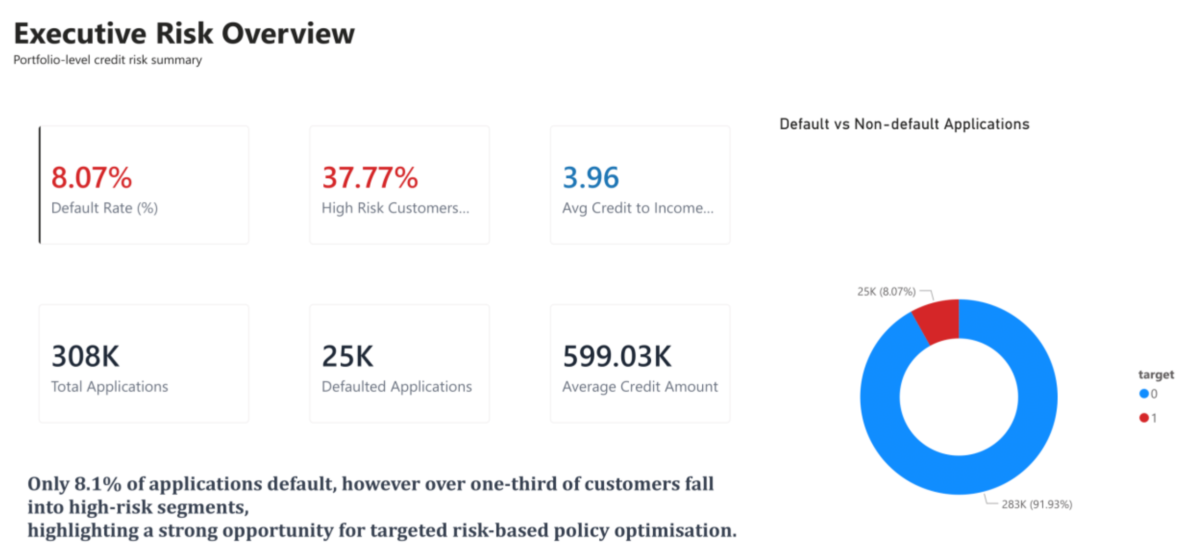
- Customer demographics (age, income)
- Credit characteristics (credit amount, credit-to-income ratio)
- Historical behaviour (previous applications, refused applications)
- Outcome variable (default indicator)

Key preprocessing steps included: - Handling missing values and invalid records - Creating categorical bands for age and income - Engineering behavioural metrics such as average previous credit and refusal counts - Validating aggregate metrics against raw source tables

This preparation ensured consistency across SQL, Python, and Power BI layers.

---

3. Executive Risk Overview (Portfolio Level)



Key Metrics

At a portfolio level, the following metrics summarise overall performance:

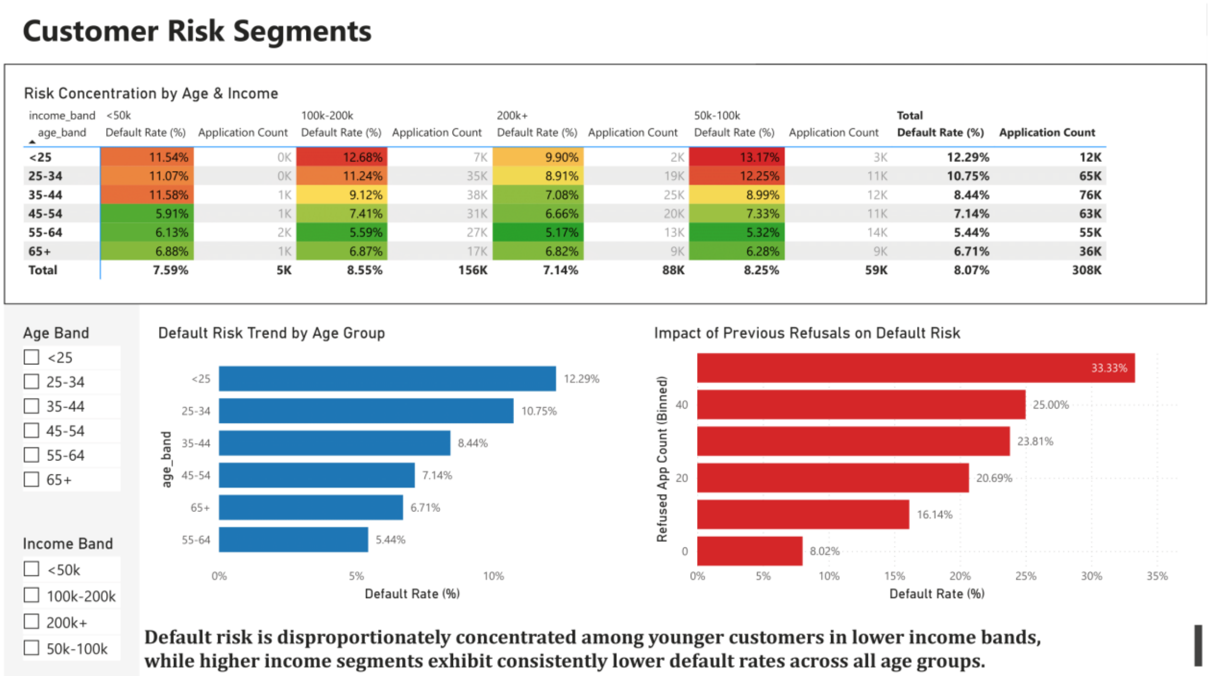
- **Default Rate:** Approximately 8% of applications result in default
- **High-Risk Customers:** Over one-third of customers fall into high-risk segments
- **Total Applications:** The portfolio contains several hundred thousand applications
- **Average Credit-to-Income Ratio:** Indicates moderate leverage across the portfolio

Insight

High-risk segments are defined based on a combination of elevated default rates and unfavourable affordability and behavioural indicators. In this analysis, customers are classified as higher risk when they belong to segments exhibiting persistently above-average default rates, high credit-to-income ratios, and/or a history of previous credit refusals.

---

4. Customer Risk Segmentation Analysis



Customer Risk Segments

Risk Concentration by Age & Income

income_band / age_band	<50k Default Rate (%)	<50k Application Count	100k-200k Default Rate (%)	100k-200k Application Count	200k+ Default Rate (%)	200k+ Application Count	50k-100k Default Rate (%)	50k-100k Application Count	Total Default Rate (%)	Total Application Count
<25	11.54%	0K	12.68%	7K	9.90%	2K	13.17%	3K	12.29%	12K
25-34	11.07%	0K	11.24%	35K	8.91%	19K	12.25%	11K	10.75%	65K
35-44	11.58%	1K	9.12%	38K	7.08%	25K	8.99%	12K	8.44%	76K
45-54	5.91%	1K	7.41%	31K	6.66%	20K	7.33%	11K	7.14%	63K
55-64	6.13%	2K	5.59%	27K	5.17%	13K	5.32%	14K	5.44%	55K
65+	6.88%	1K	6.87%	17K	6.82%	9K	6.28%	9K	6.71%	36K
Total	7.59%	5K	8.55%	156K	7.14%	88K	8.25%	59K	8.07%	308K

4.1 Risk Concentration by Age and Income

A two-dimensional risk matrix was used to analyse default rates across age bands and income bands. This matrix reveals that:

- Younger customers consistently exhibit higher default rates than older cohorts
- Lower income bands show elevated risk across nearly all age groups
- Higher income segments demonstrate relatively stable and lower default rates, regardless of age

Insight

Default risk is disproportionately concentrated among younger customers in lower income bands. This suggests that demographic and affordability factors jointly influence credit risk, and should be considered together rather than independently.

4.2 Default Risk Trends by Age Group

A ranked comparison of default rates by age group highlights a clear downward trend:

- Customers under 35 show the highest default rates
- Risk steadily declines with age
- Customers aged 55 and above consistently demonstrate the lowest default risk

Insight

Age acts as a strong proxy for financial stability and repayment behaviour. Younger customers may face income volatility, limited credit history, or higher financial pressure, increasing default likelihood.

4.3 Impact of Previous Credit Refusals

Behavioural analysis shows a strong relationship between historical refusals and future default risk:

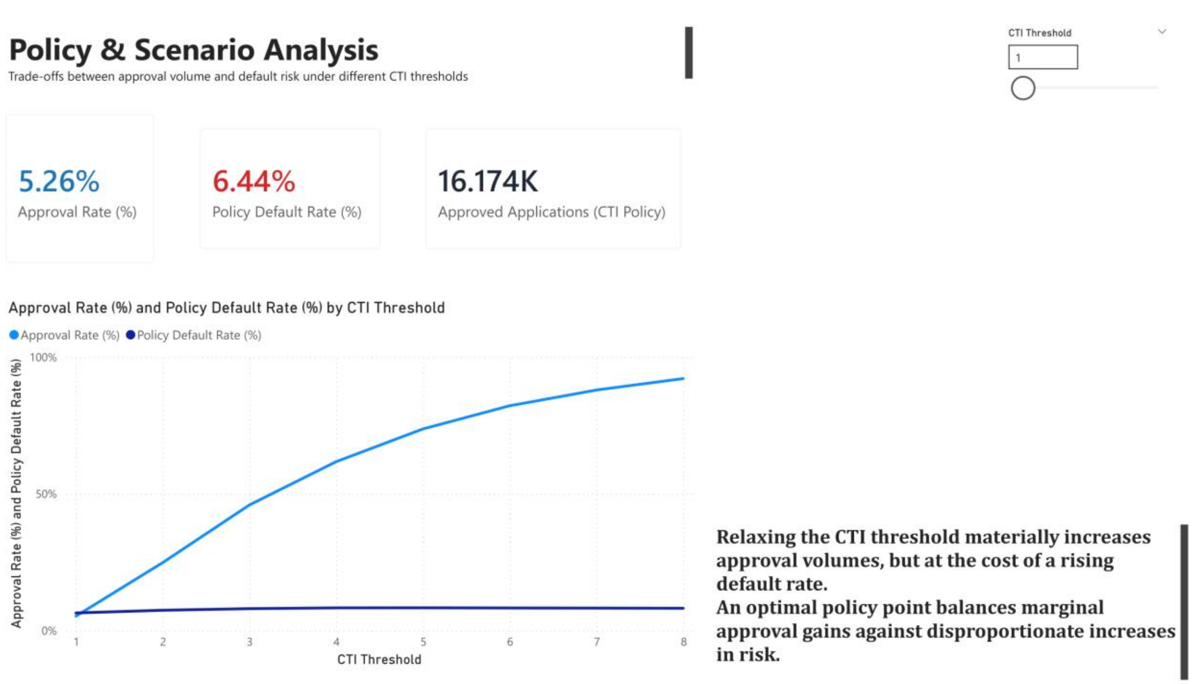
- Customers with no prior refusals exhibit the lowest default rates
- Default rates increase sharply as the number of previous refusals rises
- At high refusal counts, default risk increases disproportionately

Insight

Refused application counts were grouped into discrete behavioural bands to improve interpretability. These bands represent customers with no prior refusals, low refusal frequency, and repeated historical refusals. The analysis shows a non-linear relationship, where default risk increases modestly at low refusal levels but rises sharply once customers exceed higher refusal thresholds.

---

5. Policy and Scenario Analysis



The credit-to-income (CTI) ratio is defined as the ratio of the approved credit amount to a customer's annual income. It is used as a proxy for affordability risk, where higher CTI values indicate greater financial leverage and an increased likelihood of repayment stress.

5.1 CTI Threshold Trade-Offs

A what-if scenario analysis was conducted by varying the credit-to-income (CTI) threshold and observing its impact on:

- Approval rate
- Number of approved applications
- Policy-level default rate

The analysis shows that:

- Relaxing CTI thresholds materially increases approval volumes
- Default risk rises as approval criteria are loosened
- Beyond mid-range CTI values, approval gains begin to plateau while default risk continues to rise

Insight

There exists a clear inflection point where incremental approval gains are outweighed by disproportionate increases in default risk. This demonstrates that aggressive policy relaxation delivers diminishing returns.

---

6. Key Insights Summary

1. Portfolio-level default rates understate underlying risk concentration
2. Younger, lower-income customers represent the highest-risk segments
3. Previous credit refusals are a strong predictor of future default
4. Credit policy adjustments involve clear and measurable risk-volume trade-offs
5. A single uniform policy is suboptimal across heterogeneous customer segments

---

7. Recommendations

7.1 Segment-Based Credit Policy

Rather than applying uniform CTI thresholds, introduce differentiated policies by age and income segments. Lower-risk segments can sustain slightly relaxed thresholds, while high-risk segments require tighter controls.

7.2 Behavioural Risk Weighting

Incorporate previous refusal counts directly into credit decisioning, with non-linear penalties applied beyond defined thresholds.

7.3 Controlled Policy Optimisation

Use scenario analysis outputs to identify CTI ranges where approval gains remain efficient. Avoid threshold levels where default risk accelerates without proportional business benefit.

7.4 Ongoing Monitoring

Establish dashboards to continuously monitor segment-level default rates and policy performance, enabling early detection of risk drift.

---

8. Limitations and Next Steps

This analysis is based on historical application data and does not account for macroeconomic changes or external credit bureau updates. Future work could include:

- Time-based cohort analysis
- Predictive modelling using machine learning
- Stress testing under adverse economic scenarios

9.Project Resources & Contact

GitHub Repository
https://github.com/su7ri/home_credit

Portfolio Website
https://azed.uk

Power BI Dashboard (PBIX)
The Power BI dashboard file (PBIX) is included in the GitHub repository for local exploration and review.

LinkedIn
https://linkedin.com/in/soroush-azed

Email
soroush.azed@gmail.com

---

10. Data Source

This project uses publicly available credit risk data from Kaggle:

https://www.kaggle.com/datasets/laotse/credit-risk-dataset

The dataset contains anonymised consumer lending records, including customer demographics, financial attributes, credit amounts, and default outcomes. These variables are used to analyse default behaviour, identify high-risk segments, and evaluate credit policy trade-offs through scenario analysis.

---

11. Conclusion

This project demonstrates how integrated data analytics, segmentation analysis, and scenario modelling can support evidence-based credit policy decisions. By moving beyond aggregate metrics and focusing on segment-level behaviour and policy trade-offs, organisations can achieve a more balanced approach to growth and risk management.

Appendix A: SQL Data Preparation and Feature Engineering

This appendix documents the SQL work performed to prepare the analytical dataset used across Python and Power BI. The goal of this stage was to transform raw, multi-table transactional data into a clean, analysis-ready structure.

A.1 Data Integration (Joins)

Multiple relational tables were combined using SQL joins to create a unified customer-level view. Key joins included:

- Joining application-level data with historical credit bureau records
- Linking previous application outcomes to current applications
- Aggregating installment and credit card history at customer level

Left joins were primarily used to retain the full application population while enriching records with available historical information.

A.2 Data Cleaning and Validation

The following data quality steps were applied:

- Removal of duplicate customer records
- Handling of missing or invalid values in key financial fields
- Validation of numeric ranges (e.g. income, credit amount)
- Consistency checks between aggregated values and source tables

These steps ensured that downstream metrics were not distorted by data integrity issues.

A.3 Feature Engineering

Several analytical features were created directly in SQL to support efficient analysis:

- **Age and income bands** to enable segmentation analysis
- **Credit-to-income ratio** to capture affordability risk
- **Previous application counts** and **refusal counts** as behavioural risk indicators
- **Aggregated historical credit metrics** (e.g. average previous credit exposure)

Feature engineering in SQL reduced computational load in BI tools and ensured metric consistency across environments.

A.4 Aggregations and KPI Foundations

SQL aggregations were used to establish reliable KPI foundations, including:

- Application counts by segment
- Default counts and default rates
- Behavioural aggregates by customer group

These aggregates were validated against raw data samples before being exposed to Power BI for measure development.

A.5 Analytical Views

Reusable SQL views were created to encapsulate business logic and simplify downstream consumption. These views:

- Abstracted complex joins and transformations
- Provided a stable schema for BI and Python analysis
- Improved reproducibility and maintainability of the analytical workflow

---

A.6 Rationale for SQL-First Approach

Performing data preparation in SQL provided several advantages:

- Clear separation between data engineering and analytical layers
- Improved performance for large datasets
- Transparent, auditable transformation logic
- Easier collaboration and version control

This SQL-first approach enabled a clean transition to Python for exploratory analysis and to Power BI for interactive reporting.