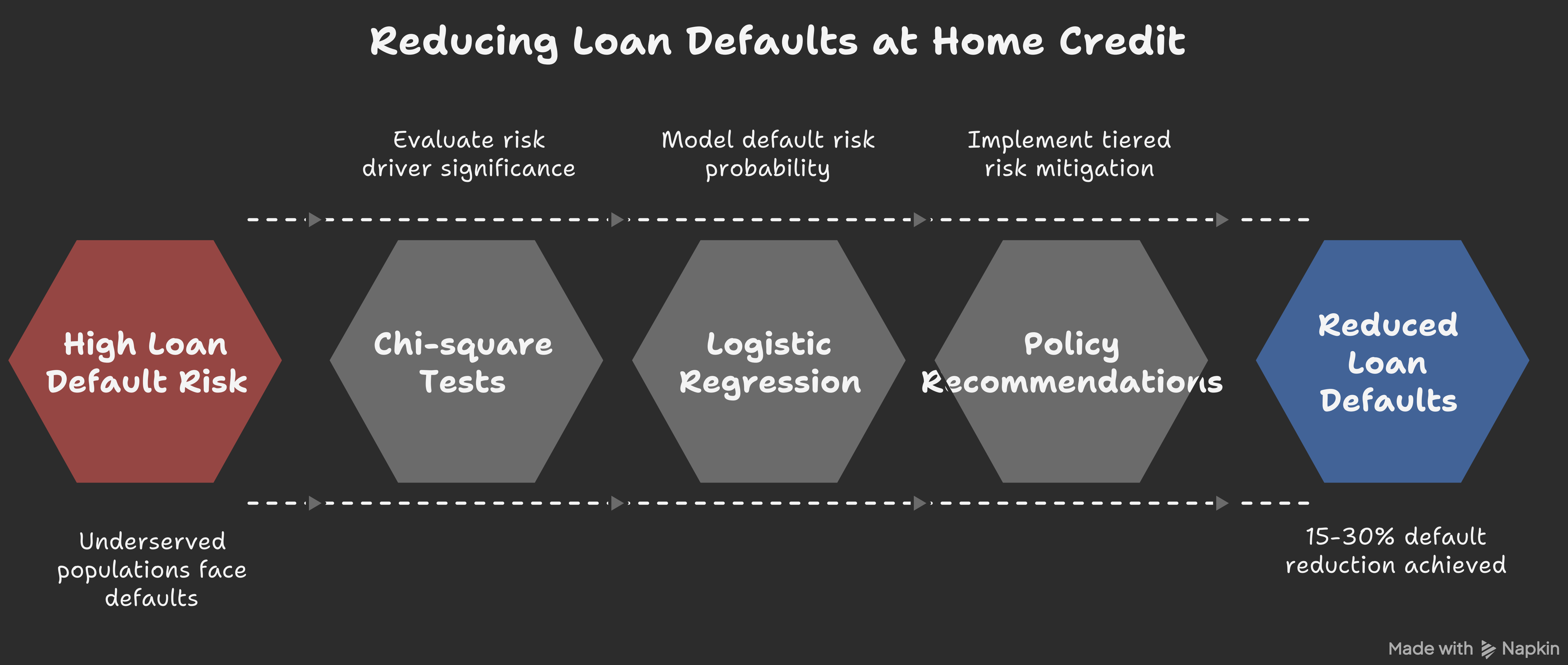
Home-Credit-Loan-Default-Risk-Analysis

Home-Credit-Loan-Default-Risk-Analysis
Analytical case study on Home Credit loan default risk: identifies key risk drivers using Chi-square tests and logistic regression. Includes actionable recommendations based on 307,511 loan applications.

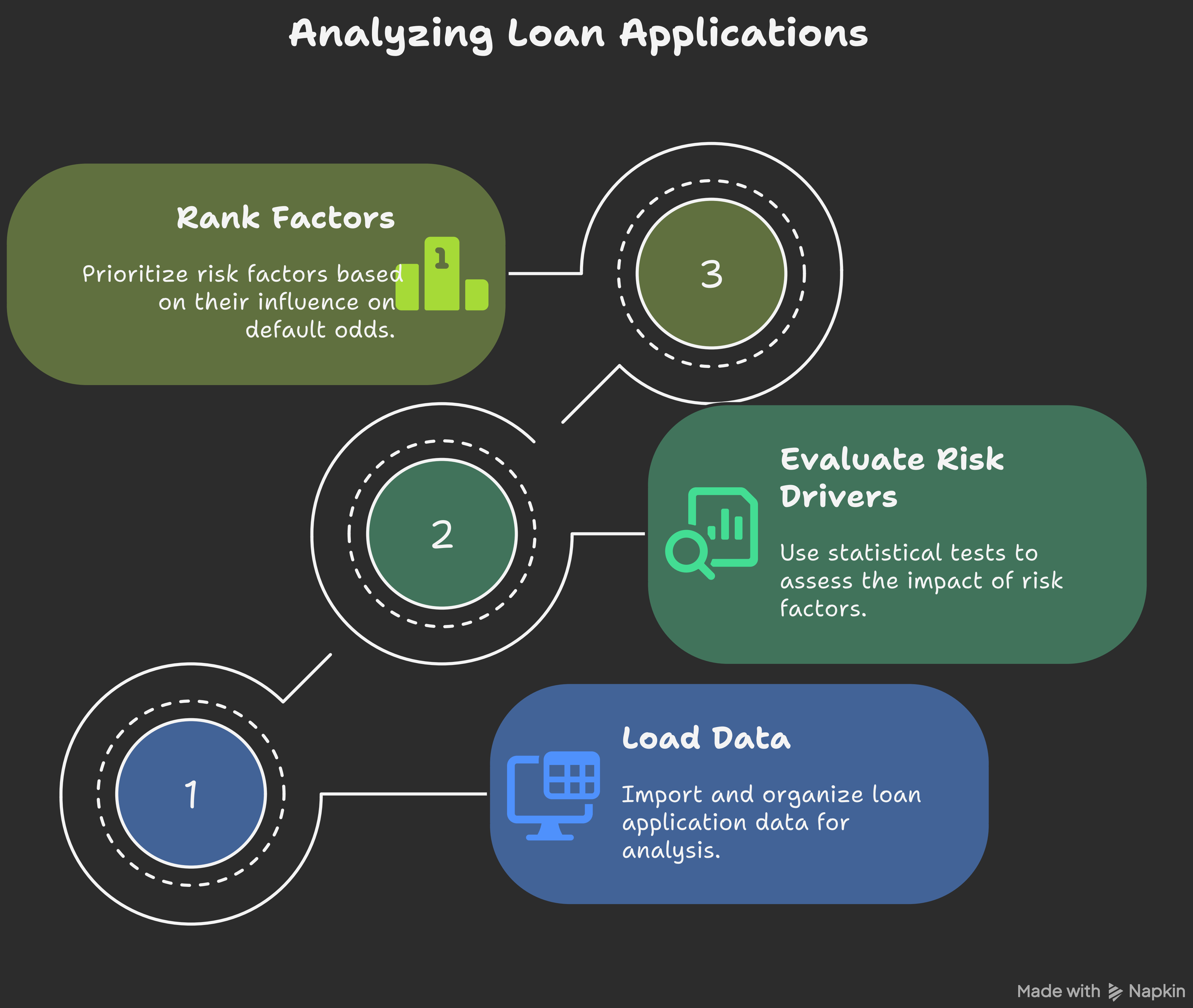
Home Credit Loan Default Risk Analysis: Case Study Workflow and Recommendations
Overview
This case study analyzes loan default risk for Home Credit, which serves financially underserved populations, using a dataset of 307,511 loan applications. The analysis leverages Chi-square tests and logistic regression to evaluate seven risk drivers: income type, regional rating, external credit scores, age cohorts, age segments, credit inquiries, and application weekdays. The objective is to provide evidence-based recommendations to reduce defaults by 15–30% while enhancing inclusion for low-risk groups.

Key objectives:
- Identify and rank impact of risk drivers on default odds.
- Propose targeted policy changes based on statistical findings.
- Estimate potential default reduction through tiered policies.



Project Structure
- **Data/**: Contains the dataset (not included here; derived from internal Home Credit records).
- **Docs/**: Includes the full case study report PDF ("case study report.pdf").
- **Visuals/**: Logistic regression results (Figures A1-A8 in the appendix).

Analytical Workflow
The project is divided into four phases:
Phase 1: Data Loading & Analysis
- Analyzed 307,511 loan applications.
- Evaluated seven risk drivers using Chi-square tests and logistic regression.
- Ranked factors by impact on default odds.



Phase 2: Key Findings - Statistical Tests
- **Income Type**: Highest impact (up to +551% odds), with unstable income sources (e.g., Unemployed +551%, Maternity Leave +61%) driving extreme risk.
- **Regional Rating**: High impact (up to +146% odds), with poorer regions (Region 3 +146%) showing higher defaults.
- **External Credit Scores**: High impact (−73.5% odds), with scores >0.5 reducing default risk significantly.
- **Age Cohorts (12 Bands)**: High impact (−61.7% odds), risk declines steadily with age.
- **Broad Age Segments**: Medium impact (−53.2% odds), seniors (>50) are safest.
- **Credit Inquiries**: Low impact (+3.7% odds), multiple inquiries slightly raise risk.
- **Application Weekday**: Lowest impact (±5.2% odds), minimal variation with Tuesday slightly riskier.

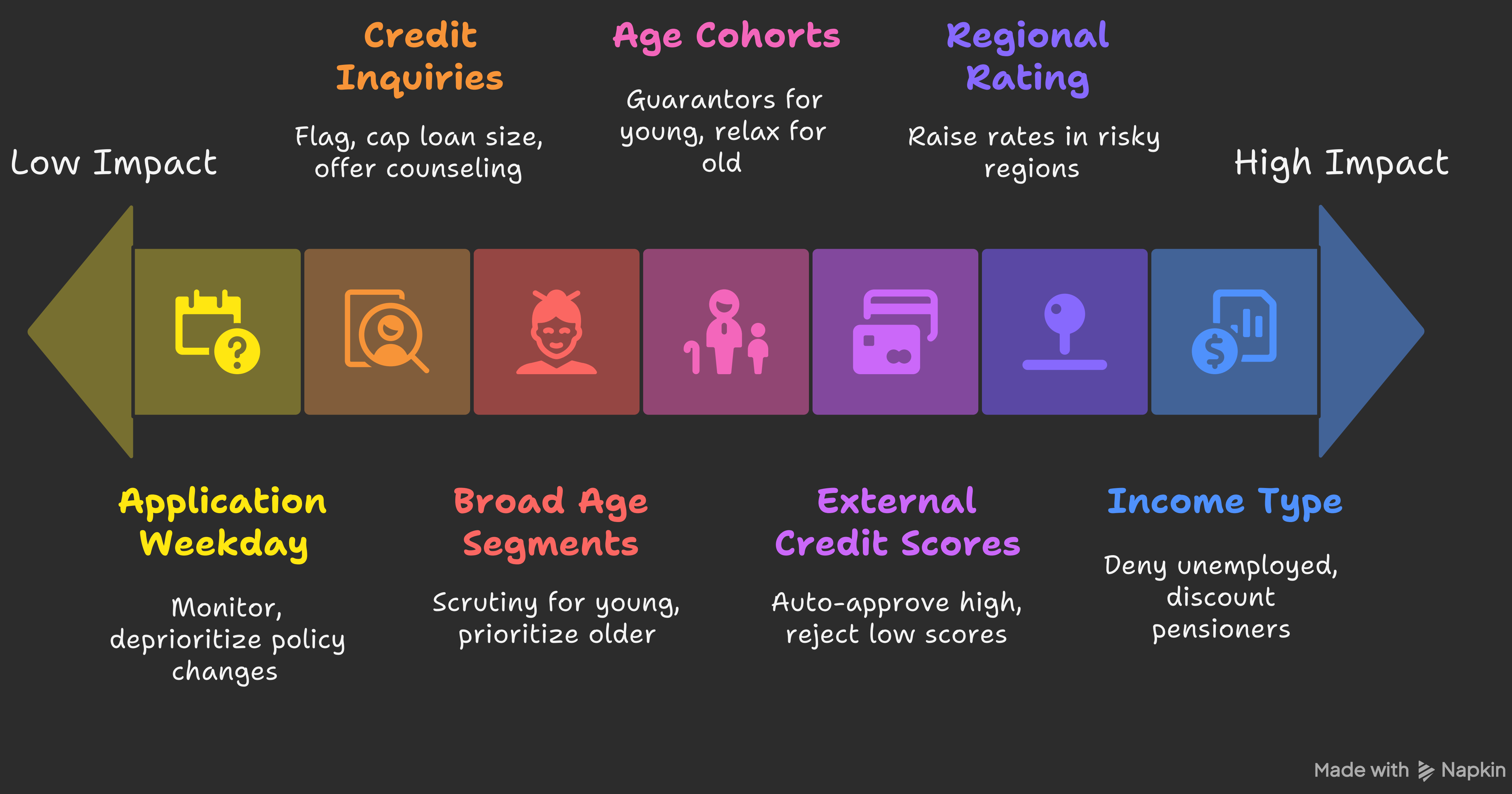
Impact of Factors on Default Risk

	Impact Level	Riskiest Group	Safest Group
Income Type	Highest	Unemployed (+551%), Maternity Leave (+61%)	Stable income sources
Regional Rating	High	Region 3 (+146%)	Wealthier regions
External Credit Scores	High	Low scores	Scores >0.5 (-73.5%)
Age Cohorts	High	Younger cohorts	Older cohorts (-61.7%)
Broad Age Segments	Medium	Younger segments	Seniors (>50) (-53.2%)
Credit Inquiries	Low	Multiple inquiries (+3.7%)	Few inquiries
Application Weekday	Lowest	Tuesday (+5.2%)	Other weekdays

Phase 3: Conclusion & Recommendations
- **Conclusion**: Top levers for risk reduction are income type, regional rating, external scores, and age.
- **Evidence-Based Recommendations**:
  - **Income Type**: Deny/cap loans for Unemployed & Maternity Leave; offer 1–2% rate discounts to Pensioners/State Servants; tighten income verification for Working/Commercial Associates.
  - **Regional Rating**: Raise rates or collateral in Region 3 by ~25%; fast-track approvals in Region 1; cap Region 3 loan amounts at 70% of standard.
  - **External Credit Scores**: Auto-approve >0.5 with minimal checks; require collateral or reject ≤0.5; expand bureau partnerships.
  - **Age Cohorts**: Require guarantors for youngest cohorts [0–3]; relax terms for oldest [9–11]; build age-based risk models.
  - **Broad Age Segments**: Extra scrutiny & financial literacy for <30; prioritize >50 in portfolio mix; limit <30 exposure to ~25%.
  - **Credit Inquiries**: Flag >2 for manual review; cap loan size; offer credit counseling.
  - **Application Weekday**: Increase Tuesday checks; promote low-risk days; monitor but deprioritize for major policy changes.

- **Expected Impact**: Tiered policies could cut defaults by 15–30% while improving inclusion for low-risk groups.

Risk reduction levers ranked by impact on loan defaults



Phase 4: Limitations & Future Work
- **Limitations**:
  - Data scope limited to provided dataset; may not capture macroeconomic shocks or post-loan behavioral changes.
  - Missing variables (e.g., debt-to-income ratio, employment tenure).
  - Model simplification assumes linear log-odds relationships.
  - Multicollinearity and lack of time-series modeling.
- **Future Work**:
  - Incorporate additional borrower metrics (e.g., payment history, debt ratios).
  - Test advanced models (XGBoost, LightGBM, survival analysis).
  - Develop multi-factor risk tiers and behavioral tracking.
  - Conduct A/B testing and enhance explainability with SHAP/partial dependence plots.

Enhancing Loan Risk Assessment: Limitations and Future Directions

Technologies Used
- Statistical tools: Chi-square tests, logistic regression.
- Software: Python-based analysis.

Getting Started
1. Review the case study report PDF ("case study report.pdf").
2. Implement recommendations based on organizational needs and data availability.

Contributing
Contributions welcome! Suggest improvements or additional analyses via feedback channels.

License
MIT License – feel free to use and adapt.

Made with 🩶!! by Gajarajan V Y. Inspired by standard dataset for educational purposes.