Credit Card Churn Data Analysis: **Relationship Depth and Engagement**

Author: Helen Bliss
Hackathon Project – Power BI Analytics

Executive Summary

By shifting from reactive retention to behaviour-led early intervention, the bank can identify risk sooner, target resources more precisely, and strengthen customer relationships before disengagement becomes irreversible.

Project Overview

This project analyses behavioural and relationship-based drivers of customer attrition using the *Bank Churners* dataset.

The output is a fully modelled, interactive Power BI dashboard designed to identify **early intervention opportunities** before disengaged customers churn.

The analysis combines:

- targeted data cleaning
- exploratory behavioural analysis
- dimensional modelling
- new custom metrics (e.g., relationship tier, engagement depth, utilisation tier, composite risk score, Z-score)
- multi-layered visual analytics

The dashboard supports business stakeholders in understanding:

1. How relationship depth affects customer stability
2. How behavioural signals (contact, inactivity, utilisation) predict churn
3. Which existing customers are currently exhibiting high-risk patterns

Role & Responsibilities

As a data analyst supporting the bank, my responsibilities included:

- Conducting independent data exploration and behavioural hypothesis generation
- Engineering segmentation features to enhance customer profiling
- Developing an analytical model of customer engagement and churn risk
- Building a Power BI dashboard with meaningful slicers, interactions, and interpretability
- Producing recommendations for preventative retention strategy
- Managing project structure, delivery, and timeline within the hackathon team

---

Hypotheses

1. Customers with deeper relationships (more products, longer tenure, higher engagement behaviour) have significantly lower churn risk.
2. Churn risk increases when disengagement indicators accumulate — specifically:
    - reduced utilisation
    - increased inactivity
    - reactive rather than proactive customer contact
3. It is possible to detect *at-risk existing customers* using behavioural composites rather than demographic variables alone.

Objectives

1. Relationship Depth & Retention

How does product ownership affect customer stability?
Objective: quantify link between product depth and churn risk.

2. Contact Behaviour & Disengagement

What is the impact of customer contact on churn?
Objective: evaluate whether contact frequency predicts risk or indicates reactive behaviour.

3. Inactivity Thresholds

   At what point does customer inactivity become dangerous?
   Objective: identify churn-critical inactivity periods for early intervention.

Methodology

1. Audit and clean the dataset for modelling readiness
2. Perform exploratory behavioural analysis to identify key drivers
3. Engineer segmentation fields to express relationship depth and engagement
4. Build a dimensional Power BI data model with lookup tables & sort orders
5. Develop calculated metrics:
   - Customer Lifecycle Stage
   - Engagement Depth
   - Utilisation Tier
   - Disengagement Indicator
   - Composite Risk Score
   - Z-Score Standardised Risk
6. Build interactive dashboard pages and slicers
7. Validate insights against business context
8. Generate recommendations for prevention-focused retention strategy

---

Data Cleaning

Data cleaning focused on establishing analytical consistency:

- Removed duplicate or invalid customer IDs
- Ensured numerical formatting for:
  - inactivity months
  - contact count
  - utilisation ratio
  - product count

- Standardised categorical variables (e.g., Income Category, Card Category)
- Validated customer tenure distributions
- Checked behavioural metric ranges using Power BI column profiling
- Created replacement values for missing income categories where appropriate
- Corrected inconsistent sort order for categorical tiers

- Investigated and resolved a major data-load issue that had limited Power BI to ~700 rows instead of 10,127. After resolution, full customer base was used for modelling

---

Exploratory Data Analysis (EDA)

Early EDA revealed expected behavioural patterns:

- Customers with **3+ products** showed significantly higher retention
- Low contact frequency was associated with stability; high contact occurred reactively among churn-prone customers
- Inactivity beyond **3 months** correlated strongly with attrition
- Declining transactional momentum (Total_Amt_Chng_Q4_Q1) signalled disengagement
- Utilisation patterns split customers into behaviourally meaningful segments

EDA was performed using:

- proportional stacked charts
- line + bar combo charts
- matrix views with conditional formatting

This stage informed which behavioural indicators should be engineered into the model.

Data Modelling (Power BI)

A dimensional schema was implemented:

Fact Table

`Table 1 (Sheet 1)` – core customer behavioural data

Dimension Tables

- Dim_Engagement_Depth
- Dim_Utilisation_Tier
- Dim_Customer_Lifecycle
- Dim_Product_Retention

These tables allowed:

- controlled sort order
- cleaner slicers
- stable categorical modelling
- prevention of circular relationship conflicts

Relationships:
One-to-many relationships were established from each dimension table to the fact table on their respective categorisation keys.

Feature Engineering: Segmentation Parameters

All segmentation thresholds were selected based on observed behavioural inflection points and domain logic, rather than arbitrary percentile splits.

Customer Lifecycle Stage

- **New Customer:** < 12 months on book
- **Growing Relationship:** 12–36 months
- **Settled Customer:** > 36 months

Rationale: Early-tenure customers show higher volatility, while churn stabilises significantly once relationships mature beyond three years.

Engagement Depth

- **Low Engagement**
- **Medium Engagement**
- **High Engagement**

Rationale: Low engagement consistently emerged as the strongest structural predictor of churn, while medium engagement formed a stable core group.

Product Depth

- **1–2 products:** Shallow relationship
- **3+ products:** Stable relationship

Rationale: Churn drops sharply once customers hold three or more products, indicating a clear stability threshold.

Utilisation Tier

- **No Utilisation:** Avg_Utilization_Ratio = 0
- **Low Utilisation:** < 0.30
- **Moderate Utilisation:** 0.30–0.70
- **High Utilisation:** > 0.70

Rationale: Churn risk is non-linear, with elevated risk at both utilisation extremes and greatest stability at moderate usage.

Inactivity Threshold

- **0–2 months:** Stable
- **3–4 months:** High-risk tipping point
- **5+ months:** Persistently elevated risk

Rationale: Churn accelerates sharply at 3–4 months of inactivity, making this the most actionable intervention window.

Composite Risk Score

- **Inputs:** Inactivity, utilisation behaviour, transactional momentum (Total_Amt_Chng_Q4_Q1)
- **Scoring:** Additive behavioural risk model

Rationale: Combines multiple early-warning signals into a single interpretable risk indicator.

Z-Score Standardised Risk

- **Metric:** (Composite Risk − Population Mean) / Standard Deviation

Rationale: Highlights customer segments that deviate most from expected behaviour, enabling prioritisation of at-risk existing customers.

Visual Analytics

[Helen Bliss: PowerBI Published Dashboard]

Tab 1 – Executive Summary

Churn by Customer Lifestage (with Disengagement Line)

Shows how churn varies across lifecycle stages, highlighting rising disengagement through an inactivity-based trend line.

Churn Rate by Engagement Depth

Compares churn across low, medium and high engagement segments, identifying structural risk among low-depth customers.

Churn by Credit Utilisation Behaviour

Reveals how utilisation patterns (none, low, moderate, high) relate to churn likelihood.

Standardised Z-Score Churn Risk Matrix ‘Composite Churn Risk by Engagement Depth and Utilisation’

Highlights which existing customer segments sit significantly above or below expected churn risk based on engagement depth and utilisation behaviour. Displays overall behavioural risk scores across customer segments.

Tab 2 – Product Retention

Customer Retention by Number of Products

Demonstrates how retention improves sharply once customers hold three or more products.

Tab 3 – Contact & Inactivity

Contact Frequency vs Inactivity

Shows how churn clusters among customers who contact the bank after becoming inactive, indicating reactive rather than preventative engagement.

Tab 4 – Inactivity & Risk

Churn by Months of Inactivity

Identifies the churn “tipping point” at 3–4 months of inactivity.

Insights

Product Ownership & Stability

Key Insight: *Churn drops dramatically once customers hold three or more products.*

Customers with only one or two products show noticeably higher attrition, but retention stabilises sharply at the **3+ product threshold**. This suggests that deeper product engagement anchors customers more strongly to the bank. The relationship is almost stepwise: each additional product reduces churn probability until stability is reached.
Why it matters: Product consolidation is one of the bank's most powerful and controllable retention levers.

Contact Behaviour

Key Insight: *It's not contact frequency that predicts churn — it's the timing.*

Attrited and existing customers contact the bank at **very similar overall levels**, meaning contact itself is not protective. Churn is highest among customers who **contact the bank after becoming inactive**, demonstrating that many interactions are **reactive rather than engagement-driven**.
Why it matters: Increasing contact volume is ineffective; the bank needs **earlier, proactive outreach** before inactivity builds.

Inactivity Threshold

Key Insight: *Churn peaks at around 4 months of inactivity — this is the bank's danger zone.*

The analysis shows stable retention up to about **2 months of inactivity**, but churn escalates sharply at **Months 3–4**, where customers shift from passive disengagement to active attrition. After 4 months, churn remains elevated but becomes harder to influence.
Why it matters: The bank should intervene **before Month 3**, when customers are still highly recoverable.

Customer Lifestage

Key Insight: *Settled customers are highly stable; new customers are not unless depth increases quickly.*

Settled customers show consistently low churn, reflecting long-term behavioural commitment. New customers, however, show **volatile churn patterns**, especially when product depth and engagement remain low. Growing customers sit in the middle, improving as engagement deepens.

Why it matters: Early relationship development is essential; onboarding quality and product adoption in the first months make a disproportionate difference.

Engagement Depth

Key Insight: *Low engagement is the strongest structural predictor of churn.*

Across every segment examined, low-engagement customers show a disproportionately high churn rate. Medium engagement forms the stable "core" of the customer base, while high engagement offers incremental retention benefit but less dramatic uplift than moving someone out of low engagement.

Why it matters: Improving engagement depth — even modestly — is one of the highest-impact retention strategies.

Utilisation Behaviour

Key Insight: *Both extremes — no utilisation and very high utilisation — carry elevated churn risk.*

Customers who **never use their card** behave similarly to inactive customers, signalling a weak attachment to the product. Conversely, customers with **very high utilisation** often show stress behaviour or credit fatigue, both associated with churn. **Moderate utilisation** is consistently the most stable category.

Why it matters: Utilisation is not linear — it produces **two risk tails**, each requiring different interventions.

Composite Risk Heatmap

Key Insight: *The risk distribution forms a behavioural bell curve — with two clear risk tails.*

When churn risk is standardised using a Z-score, **medium engagement customers form the expected centre**, showing typical behaviour with no elevated risk. Low-engagement customers, however, **split into two high-risk tails**:

- those with **no utilisation**, and
- those with **high utilisation**.

Filtering the heatmap to **existing customers only** reveals which live segments currently deviate most from expected behaviour and therefore require immediate attention.

Why it matters: This is a powerful diagnostic tool, showing precisely where to focus resources for maximum retention impact.

Hypothesis Conclusions

Hypothesis 1

Customers with deeper relationships (more products, longer tenure, higher engagement) have significantly lower churn risk.

Conclusion:
This hypothesis is **strongly supported**. Customers holding **three or more products** show a sharp and sustained reduction in churn, while those classified as **Settled** and **Medium–High Engagement** consistently demonstrate the lowest risk across all analyses. Relationship depth acts as a stabilising force, with product ownership emerging as one of the most powerful and controllable retention levers available to the bank.

Hypothesis 2

Churn risk increases when disengagement indicators accumulate, particularly reduced utilisation, increased inactivity, and reactive contact behaviour.

Conclusion:
This hypothesis is **confirmed**. Churn does not correlate with contact frequency alone, but instead rises when customers become inactive and subsequently contact the bank, indicating **reactive rather than preventative engagement**. Risk escalates further when inactivity is combined with utilisation extremes—either no usage or very high usage—demonstrating that disengagement is multi-dimensional and accumulative rather than driven by a single factor.

Hypothesis 3

At-risk existing customers can be identified using behavioural composites rather than demographic variables alone.

Conclusion:
This hypothesis is **fully validated**. The Composite Risk Score and Z-score standardisation successfully identify existing customer segments that deviate significantly from expected behaviour, particularly among low-engagement groups. These behavioural composites outperform single demographic or transactional measures by capturing multiple early-warning signals simultaneously, enabling precise and actionable risk identification before churn occurs.

Overall Conclusion

Together, these findings demonstrate that customer churn is best understood as a **behavioural progression**, not a sudden event. By focusing on relationship depth, engagement behaviour, and early disengagement signals, the bank can move from reactive churn management to proactive, prevention-led retention strategies.

Business Recommendations

1. Deepen Relationships Early

Insight link: Churn drops sharply once customers hold **3+ products**.

Promote early multi-product adoption, particularly focusing on the transition from **1–2 products to 3 products**, where stability increases most dramatically.
Target this within the first months of the relationship, when customers are most volatile.

Operational actions:

- Embed product recommendations into onboarding journeys
- Offer bundled incentives for second and third products
- Prioritise cross-sell for low-engagement new customers

2. Monitor Early Inactivity (Prevent the Danger Zone)

Insight link: Churn accelerates at **3–4 months of inactivity**.

Introduce proactive monitoring and soft interventions at **2 months of inactivity**, before customers enter the high-risk period.

Operational actions:

- Automated reminders or usage nudges
- Light-touch "check-in" messaging
- Friction-reduction prompts (e.g. payment setup, digital wallet usage)

3. Reframe Contact Strategy (From Reactive to Preventative)

Insight link: Contact is common among churners, but occurs **too late** to prevent attrition.

Shift the focus away from volume-based contact metrics and towards **timely, behaviour-triggered engagement**.

Operational actions:

- Trigger outreach based on inactivity or declining spend, not complaints
- Use behavioural thresholds (e.g. no usage in 45–60 days) to prompt action
- Personalise messaging around usage encouragement rather than issue resolution

4. Prioritise High-Risk Low-Engagement Segments

Insight link: Low engagement combined with utilisation extremes forms two distinct risk tails.

Use the composite risk heatmap to prioritise outreach to customers showing:

- **Low engagement + no utilisation** (early disengagement)
- **Low engagement + high utilisation** (stress or credit fatigue)

Operational actions:

- Differentiate messaging for each risk type
- Support low-use customers with activation guidance
- Support high-use customers with budgeting tools or credit reassurance

5. Build a Relationship Growth Pathway

Insight link: Settled customers are consistently the most stable.

Operationalise the engineered **Relationship Tiers** (New → Growing → Settled) as a structured customer lifecycle framework.

Operational actions:

- Define success criteria for moving between tiers
- Align communications and offers to lifecycle stage
  Measure success by progression rate, not just churn reduction

6. Introduce Behaviour-Based Risk Scoring (Next-Step Enhancement)

Insight link: Composite and Z-score risk measures clearly identify deviation from expected behaviour.

Integrate behavioural risk scores into CRM and reporting systems to support **early-warning detection** and prioritisation.

Operational actions:

- Flag customers with elevated Z-score risk for review
- Combine risk score with lifecycle stage for targeted intervention

- Use the score to test and refine retention strategies over time

7. Measure Success Using Behavioural Change, Not Just Retention

Insight link: Engagement depth and utilisation patterns change before churn occurs.

Track intermediate success metrics such as:

- Reduction in inactivity duration
- Movement from low → medium engagement
- Increase in product depth
- Stabilisation of utilisation behaviour

This ensures interventions are improving **behaviour**, not simply delaying churn.

Future Development

Building on the behavioural segmentation and dashboard insights developed in this project, several opportunities exist to extend the analysis into more advanced and operationally meaningful directions.

1. Predictive Modelling Using a Train/Test Split

The next logical step would be to evolve the descriptive analysis into a fully predictive churn model. This would involve exporting the cleaned dataset and developing a Python-based machine learning pipeline using an 80/20 train/test split. The engineered behavioural features created during this project—such as engagement depth, utilisation tier, inactivity indicators and the composite risk score—would serve as strong predictors for a classification model.

The modelling workflow would include:

- Encoding categorical variables.
- Establishing a baseline model using Logistic Regression for interpretability.
- Training advanced models such as Random Forest or Gradient Boosting (e.g., XGBoost or LightGBM) to improve predictive performance.
- Evaluating accuracy, precision, recall, ROC-AUC, and the confusion matrix to assess model quality.
- Conducting feature importance analysis to validate which behaviours most strongly influence churn risk.

The predicted churn probability generated by the model could then be re-imported into Power BI as a new column, allowing the dashboard to display both observed and predicted risk for each customer segment.

This would transform the dashboard into a proactive early-warning system rather than a retrospective analysis tool.

2. Next Best Product Recommendation Engine

Given the strong relationship identified between product depth and customer stability, there is significant value in developing a “next best product” recommendation model. Using historical adoption patterns, the model could suggest which additional product each customer is most likely to adopt. This would help the bank prioritise personalised cross-sell strategies for customers with low product depth—one of the clearest retention levers identified in the analysis.

3. Behaviour Drift Monitoring

A further enhancement would involve tracking customer behaviour over time rather than relying solely on static metrics. Incorporating longitudinal trends such as changes in utilisation, inactivity, and spending momentum would allow the bank to detect early signs of disengagement even before the customer enters a high-risk category. This time-series approach would also enhance the predictive model’s performance.

4. Automated Retention Trigger System

Finally, the outputs of the dashboard and predictive model could be incorporated into the bank’s CRM systems to create automated intervention triggers. Customers displaying:

- rapid increases in inactivity,
- extreme utilisation changes, or
  high model-predicted churn probability

could be flagged for immediate retention action. This would close the loop between analytics and operations, ensuring that insights translate into meaningful business outcomes.