Thursday, 17 July 2025
GenAI Powered Data Analytics
- by Tata Insights and Tata iQ

Hi,my name is deep and in this pdf I will share how I did this the tasks assigned to me and things to know about this course.

Subject
- Assigned as a role of AI Transformation Consultant within the financial services sector
- Solving a real world business problem : identifying and preventing credit card deliquescy using the power of gen ai and advanced analytics
- high credit card delinquency rate" means : A large percentage of credit card users are failing to make their payments on time, especially minimum dues.
- Goals : primary objective is to analyze customer data and predict delinquency risks using AI-driven techniques.
- My responsibilities : analyzing customer data, developing predictive models to assess delinquency risk, and recommending AI-driven solutions to improve financial decision-making.
- Project with geldium, a financial services provider specialising in digital leading and consumer credit card
- Increased in credit card delinquency rates
- Task :
  - Review available customer data to identify patterns and risk factors.
  - Develop a predictive model logic using GenAI to forecast delinquency likelihood.
  - Provide structured recommendations on intervention strategies.
  - Prepare a PowerPoint presentation outlining how Geldium can implement an AI-powered collections system, including ethical guardrails and expected business impact.
- Task 1 : to uncover patterns, risks, opportunities hidden in financial data…will use genAI …and start by improving data qualities…will use synthetic data generation techniques..used when real world data in incomplete…
  - First stage : first stage of the project: conducting an exploratory data analysis (EDA) to assess dataset completeness, identify patterns, and flag potential gaps that could impact delinquency predictions.
  - Identifying missing data or inconsistent datapoints

- Using genAI to generate insights ensuring data confidentiality and avoiding the exposure of sensitive financial information.
- Summarise the patterns, anomalies, and risk indicators that should be considered in later stages of the project in a report.
- EDA(Exploratory data analysis) : first step of understanding dataset…helps uncover patterns, trends, inconsistencies, and missing values…ensures data integrity, highlight anomalies, prevents biased models (unfairly favour or discriminate)

Four steps :

1. Understanding the Dataset
    - Identify key variables (e.g., income, credit utilization).
    - Check for categorical/numerical columns, missing/inconsistent data.
    - GenAI Use: Summarise dataset, column meanings, data types, patterns.

    Prompt: "Summarise key columns and missing values in this dataset."

2. Handling Missing Values & Outliers
    - Use mean/median or regression for statistical imputation.
    - Determine missingness type: MCAR, MAR, MNAR.
    - Drop irrelevant features only if necessary.
    - GenAI Use: Detect missing data, suggest imputation strategies.

    Prompt: "Identify missing data and suggest best imputation method."

3. Exploring Relationships Between Variables
    - Analyse how features (like income, age, credit score) affect delinquency.
    - Look for correlations or trends.
    - GenAI Use: Generate correlation matrix, explain insights in plain language.

    Prompt: "Show correlation between income and late payments."

4. Detecting Patterns & Risk Factors
    - Spot behavioural trends (e.g., frequent missed payments).
    - Group high-risk profiles (e.g., new accounts, low-income).
    - GenAI Use: Highlight patterns and risk indicators.

    Prompt: "Find top 3 risk factors for credit delinquency."

Causes of Missing Data

- Random Errors: e.g., system glitches.

- Skewed Collection: e.g., some customers avoid reporting income.
- Customer Behavior: e.g., hiding debt due to financial distress.

Techniques to Handle Missing Values

1. Deletion (Remove Rows/Columns)
   - Use only if few values are missing.
   - Risk: Reduces data size.

2. Imputation (Fill Missing Values)
   - Mean/Median/Mode Imputation: Replace missing with average values.
   - Forward/Backward Fill: Use previous/next values (common in time series).
   - Regression Imputation: Predict missing values using other variables.

3. GenAI-Assisted Imputation
   - Use AI to:
   - Detect missing value patterns.
   - Suggest appropriate imputation.
   - (Optionally) Generate synthetic data, if needed, with care.

Note: Don't blindly trust GenAI suggestions — always validate with domain knowledge to avoid bias.

Other Data Quality Checks
- Remove Duplicates
- Fix Format Inconsistencies
- Check Logical Errors

e.g., high credit score with recent payment defaults.

Bottom Line

Combine statistical techniques with AI tools for clean, unbiased, and trustworthy data — critical for making accurate predictions in financial risk analysis.

- customer risk factors for credit card delinquency
- Key risk factors for delinquency : (i) Payment History (ii) credit utilisation rate (iii)DTI ration (iv) recent credit card activity (v) employment and income stability (vi) demographics

- AI / ML angle : evuluate multiple risks at a tym, detect hidden patterns.
- What is Synthetic Data?
  - Artificially generated data that mimics real-world patterns
  - Created using statistical models or AI techniques
  - Helps when data is missing, sensitive, or limited
- When to Use Synthetic Data
  1. Enhancing Small Datasets
     → Supplement training data when actual records are limited
  2. Filling Missing Values
     → Approximate realistic values when key data is missing
  3. Testing AI Models
     → Simulate scenarios (e.g., delinquency spikes) before deployment
  4. Ensuring Privacy Compliance
     → No real customer data exposed
- Best Practices
  - Validate against real-world distributions
  - Use GenAI + traditional models (e.g., Monte Carlo, bootstrapping)
  - Never use synthetic data as the primary source in regulated settings
- Example GenAI Prompt

"Generate synthetic payment history data for customers with missing records while ensuring that distributions align with historical patterns (e.g., standard deviations, default trends)."

- Ethical Considerations
  - Avoid bias and unrealistic patterns
  - Preserve fairness and regulatory compliance
  - Ensure transparency and reproducibility in generation methods
- FROM here my tasks starts …made a EDA report

- TASK 2 : using AI detect which customers will miss there credit card payment ..to use the most suitable approach for model
- GenAI for Predictive Modeling (No Coding)

  What it is:

  Predictive modeling forecasts outcomes (like loan defaults) using past data.

  How GenAI helps:

  - Suggests models (e.g., logistic regression) based on your data
  - Explains logic in simple language

- Interprets results (e.g., why a customer might default)
- Generates code (optional) or lets you build models using prompts
- Improves models with follow-up questions (no coding needed)

Use prompts like:

- “Which features best predict delinquency?”
- “Explain this model’s prediction in simple terms.”

Important: Always verify outputs for fairness, accuracy, and bias before using in business decisions.

- How GenAI Assists in Predictive Modeling (Quick Guide)

1. Model Selection
   - GenAI recommends the best models based on your dataset:
   - Decision Trees → Easy to explain
   - Logistic Regression → Great for yes/no (delinquent or not)
   - Neural Networks → Handles complex data (less explainable)

“Which model suits this credit risk dataset and why?”

2. No-Code Model Generation
   - GenAI can auto-generate Python, R, or SQL code for models like logistic regression.
   - Acts as a starter template, but manual validation is a must.

Prompt:

“Write a logistic regression model to predict credit card delinquency, and explain each step.”

3. Model Evaluation & Fairness
   - GenAI helps:
   - Suggest metrics: accuracy, precision, recall
   - Interpret results
   - Detect bias (e.g., against certain age groups or income levels)

Prompt:

“Evaluate this model’s recall score and highlight any potential biases in predictions.”

Bottom line: GenAI simplifies modeling for non-coders but results must be verified by experts for fairness and reliability.

- Common Models for Credit Risk Assessment

1. Decision Trees

   Easy to interpret

   Works with both numbers & categories

   Shows which features drive risk

   Best when: You need transparency & explainable risk rules

Prompt:

“Build a decision tree using income & missed payments to classify delinquency risk.

2. Logistic Regression

Predicts probability (0–1) of delinquency

Easy to explain to stakeholders

Works well with structured data

Best when: You want a simple, reliable risk score

Prompt:

“Create a logistic model using income, DTI, and payment history to predict default.”

3. Neural Networks

Detects complex, deep patterns

Best for large, detailed datasets

Less explainable (“black box”)

Best when: You need high accuracy & can handle less transparency

Prompt:

“Build a basic neural network for credit risk prediction and compare it to logistic regression.”

- Evaluating Model Performance

Key Metrics:

- Accuracy – Overall correctness.
- Precision – How many predicted delinquents are actually delinquent.
- Recall – How many real delinquents were correctly identified.
- F1 Score – Balance between precision & recall.
- AUC-ROC – How well the model separates delinquent vs. non-delinquent.
- Confusion Matrix – Visual of prediction vs. actual.

If Performance is Poor:

- Improve features (e.g. add customer tenure).
- Balance the dataset (handle class imbalance).
- Try different models (tree vs. logistic).
- Tune model parameters.

Bias, Explainability & Fairness

Bias:

Models may unintentionally discriminate based on features like income, ZIP code, etc.

Use fairness metrics (e.g., demographic parity) to check bias.

Explainability:

- Decision Trees & Logistic Regression = transparent
- Neural Networks = harder to explain
- Use SHAP for detailed explanation.

Fairness:

- Avoid unfair treatment of certain groups.
- Use diverse data & fairness testing tools.

Using GenAI in Predictive Modeling

GenAI Can Help With:

- Model selection suggestions.
- Generating initial model code (Python/R/SQL).
- Preprocessing and feature selection.
- Tuning hyperparameters.
- Performance evaluation suggestions.

Prompt:

“Generate a logistic regression model to predict credit card delinquency with preprocessing and evaluation.”

GenAI simplifies the process, but human oversight is essential to ensure models are accurate, fair, and compliant.

Turning Predictive Insights into Business Value for Geldium

Step 1: Restate the Insight

Our predictive model indicates that:

- Customers under 30 years old
- With high credit utilization (>75%)
- And two or more missed payments in the past 6 months

are 3.4 times more likely to become delinquent compared to the average customer.

Step 2: Brainstorm Actions

1. Send proactive SMS/email payment reminders to high-risk segment.
2. Launch a 6-week financial coaching program for Gen Z customers.
3. Offer temporary hardship relief options.
4. Lower credit limits temporarily for at-risk customers.
5. Provide personalized repayment plan options.
6. Introduce a chatbot for real-time support on billing and repayments.

Step 3: Evaluate Options

Step 4: Form the Recommendation

SMART Recommendation: "Launch a targeted 6-week financial coaching and communication campaign for Gen Z customers (under 30) with high credit utilization and missed payments. The goal is to reduce 30+ day delinquency by at least 12% within this cohort."

Step 5: Justify It

- Scalability: Can be rolled out in phases and expanded to other cohorts.
- Ease of Implementation: Leverages existing communication and education platforms.
- Fairness: Focuses on education and empowerment, not restriction.

- Anticipated ROI: A 12% reduction in delinquency can significantly reduce losses.
- Business Alignment: Supports Geldium's goals of proactive risk management, customer retention, and ethical collections.
- Evidence-based: Based directly on model outputs and validated through risk scoring.

Presentation Strategy:

- Lead with: "Customers under 30 with high utilization and missed payments are 3.4x more likely to go delinquent."
- Show: Clear charts (e.g., bar graph of risk by age/utilization).
- Recommend: Coaching + communication = informed, empowered customers.
- Reassure: Actionable, fair, measurable, aligned with brand.

Ethical Guardrails Embedded:

- Bias Checks: Model performance evaluated across age, income, gender.
- Explainability: Model uses interpretable decision tree logic.
- Responsible Use: No punitive action taken; only supportive, educational outreach.

By embedding fairness, transparency, and strategic impact into our AI recommendations, Geldium can make data-driven decisions that reduce risk and strengthen customer trust.

Title: AI-Powered Collections System Framework for Geldium

Slide 1: Title Slide

Title: Designing a Smart, Adaptive & Ethical AI Collections System for Geldium

Subtitle: Moving from Risk Identification to Autonomous Action

Presented by: [Your Name] | Tata iQ Analytics Team

Slide 2: Problem Statement & Opportunity

- Current process: Manual outreach and static reports
- Business need: Real-time, automated decision-making at scale
- Vision: Agentic AI system that autonomously manages delinquency risk with fairness and responsibility

---

Slide 3: System Architecture Overview
Four Core Components:

1. Data Pipeline – Ingests real-time data (e.g., risk scores, demographics, payment activity)
2. Decision Engine – Applies business logic + AI models to choose best action
3. Action Layer – Executes personalized interventions (SMS, chatbot, coaching offers)
4. Learning Loop – Monitors outcomes and retrains/adjusts strategy

Diagram: Flowchart showing Data > Engine > Action > Feedback

Slide 4: Agentic AI in Action
Agentic AI capabilities:

- Autonomously chooses outreach based on evolving customer behavior
- Learns from outcomes and adjusts in real time
- Balances business goals (reduce risk) and customer experience (supportive)

Example: A customer misses 2 payments. Instead of fixed messaging, AI offers deferral + coaching based on income volatility pattern.

Slide 5: Human Oversight & Boundaries
When humans must intervene:

- Denial of hardship or escalation to collections
- Edge cases or anomalies flagged by the system
- Quarterly reviews of model decisions and outcomes

Controls:

- Human-in-loop for high-impact decisions
- Manual override options
- Escalation workflows built in

Slide 6: Fairness, Transparency, & Compliance
Key Safeguards:

- Explainability: Use interpretable models (decision trees, SHAP for black-box models)
- Bias Checks: Audit by age, income, gender (disparate impact tests)
- Data Privacy: Compliant with GDPR, FCRA, ECOA
- Audit Trails: Logged decisions with rationale and inputs

Slide 7: Regulatory Alignment Strategy

- Traceable decision flow with documentation
- Compliance integrated early (risk/legal teams)
- Manual reviews for flagged predictions
- Continuous monitoring of model fairness + accuracy
- Adaptable system as laws evolve

Slide 8: Roadmap to Implementation
Phase 1: Pilot AI decision engine with high-risk segment
Phase 2: Automate outreach + feedback collection
Phase 3: Integrate full learning loop with periodic audits
Phase 4: Scale to all segments with live compliance dashboard

---

Slide 9: Key Benefits

- Reduction in 30+ day delinquency (target: 12% in 6 months)
- Personalized outreach at scale
- Transparent, fair, and compliant decisions
- Enhanced customer trust and operational efficiency

Slide 10: Conclusion & Next Steps

- Build agentic, responsible AI into collections strategy

- Balance autonomy with oversight
- Deliver measurable outcomes ethically

Next Step: Stakeholder workshop to align implementation roadmap & success metrics

Tata Group Data Analytics Job Simulation on Forage - July 2025

- Completed a job simulation involving AI-powered data analytics and strategy development for the Financial Services team at Tata iQ.
- Conducted exploratory data analysis (EDA) using GenAI tools to assess data quality, identify risk indicators, and structure insights for predictive modeling.
- Proposed and justified an initial no-code predictive modeling framework to assess customer delinquency risk, leveraging GenAI for structured model logic and evaluation criteria.
- Designed an AI-driven collections strategy leveraging agentic AI and automation, incorporating ethical AI principles, regulatory compliance, and scalable implementation frameworks.

“Why are you interested in this role?”
I recently participated in Tata iQ's job simulation on the Forage platform, and it was incredibly useful to understand what it might be like to be on a data analytics team in an AI transformation consulting role.
I worked on a project to develop an AI-powered strategy for predicting and reducing customer credit delinquency in financial services. I practiced using GenAI tools like ChatGPT and Google Gemini and built my data analysis, predictive modeling, and responsible AI strategy skills in a real-world context.
Doing this program confirmed that I really enjoy working on data-driven business problems that combine analytics with practical strategy and I'm excited to apply these skills on a data analytics team at a company like Tata iQ.