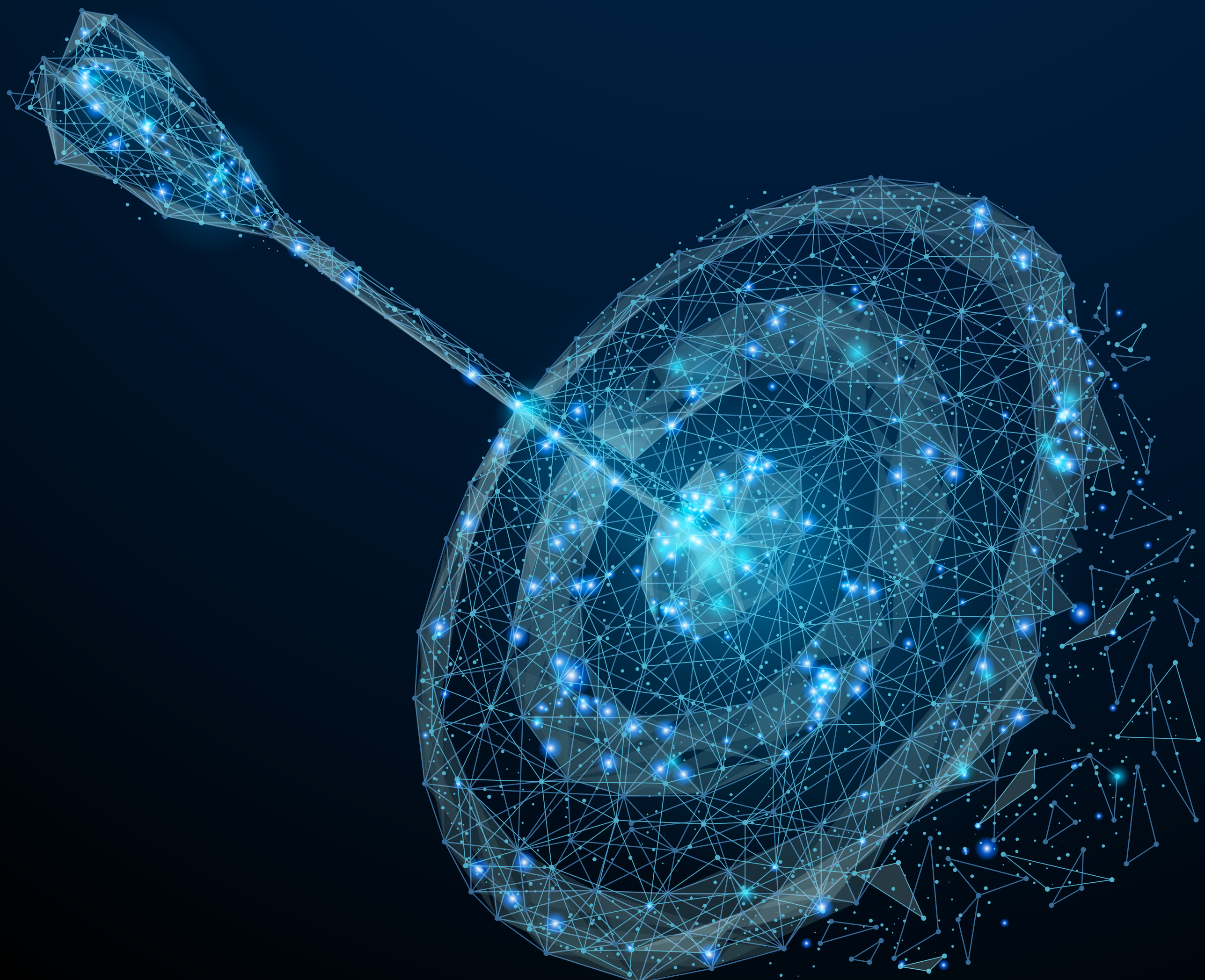
DMS DIGITAL MATRIX SYSTEMS

Monitoring Borrower Data:

How to Leverage Advanced Monitoring Techniques and Improve Results



Introduction

Accuracy and consistency of borrower data are critical for making sound lending decisions and managing risk. The recent shift by major credit bureaus to cloud-based platforms marks a significant evolution in how data is managed and delivered. This transition promises greater performance, scalability, and flexibility in data processing and delivery.

However, these technological advancements also pose challenges—particularly in maintaining the continuity and reliability of predictive attribute values. Data integrity during migrations is essential, as even minor discrepancies can lead to significant impacts on credit decisions, risk assessments, and regulatory compliance.

This paper delves into advanced strategies for monitoring changes in borrower data, including those that may result from transitions to cloud-based platforms. We'll explore a comprehensive toolkit of monitoring techniques, including Attribute Control Charts (ACCs), Population Stability Index (PSI) assessments, Characteristic Analysis Reports, and Scorecard Performance Reports. These methods will not only help you identify and understand shifts in data but also ensure that your credit decision-making processes remain robust and accurate.

Understanding Attribute Control Charts (ACCs)

Attribute Control Charts are invaluable for visualizing and monitoring variations in categorical data over time. For financial institutions, ACCs can track shifts in borrower attributes that are essential for risk assessment, such as credit inquiries or delinquency counts. These charts help differentiate between normal variations and significant deviations that warrant further investigation, thereby safeguarding the integrity of your credit scoring models.

How ACCs Work

ACCs plot the values of a specific attribute over time, typically including:

1. A center line representing the average value
2. Upper and lower control limits, usually set at ±3 standard deviations from the mean
3. Data points representing the attribute values at different time intervals

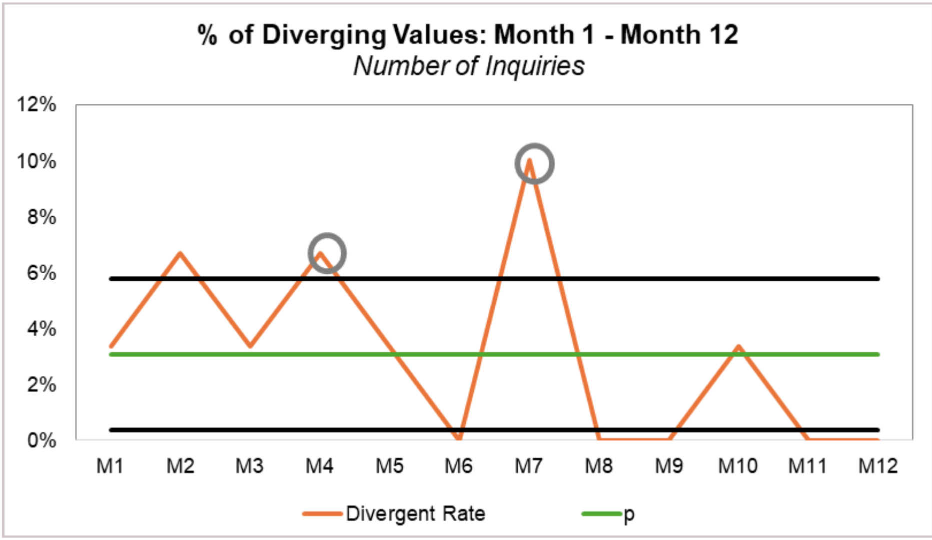
Example of an ACC

Imagine tracking the number of credit inquiries a borrower has made within the last twelve months:

*'Diverging values' fall outside of: "Mean +/-(3*Stdev)"

Month	Sample Size	Diverging Values	p	Divergent Rate	Lower Limit	Upper Limit
M1	30	1	3.056%	3.333%	0.33%	5.78%
M2	30	2	3.056%	6.667%	0.33%	5.78%
M3	30	1	3.056%	3.333%	0.33%	5.78%
M4	30	2	3.056%	6.667%	0.33%	5.78%
M5	30	1	3.056%	3.333%	0.33%	5.78%
M6	30	0	3.056%	0.000%	0.33%	5.78%
M7	30	3	3.056%	10.000%	0.33%	5.78%
M8	30	0	3.056%	0.000%	0.33%	5.78%
M9	30	0	3.056%	0.000%	0.33%	5.78%
M10	30	1	3.056%	3.333%	0.33%	5.78%
M11	30	0	3.056%	0.000%	0.33%	5.78%
M12	30	0	3.056%	0.000%	0.33%	5.78%
Total	360	11				

In this visual representation, any points falling outside the control limits would warrant investigation, as they may indicate significant shifts in inquiry patterns.



Limitations of ACCs

While ACCs are powerful, they have limitations:

1. They focus on individual attributes and may miss broader population shifts
2. They don't provide insight into the impact of changes on model performance
3. They may be sensitive to outliers or seasonal variations

These limitations underscore the need for complementary monitoring techniques.

Detecting Shifts in Data

Several factors can cause shifts in borrower data:

1. **Changes in Marketing Campaigns:** New marketing strategies might attract different customer segments, leading to changes in applicant attributes. For example, a campaign targeting young professionals could shift the age distribution of your applicant pool.
2. **Population Shifts:** Economic and demographic changes can alter the characteristics of your borrower pool. A recession might increase the proportion of applicants with recent delinquencies, while gentrification in certain areas could change the income distribution of mortgage applicants.
3. **Data Migration Effects:** The transition to cloud infrastructure can inadvertently introduce changes or errors in how data attributes are reported and interpreted. For instance, the way credit inquiries are counted or categorized might change, leading to apparent shifts in inquiry patterns.

The potential impact of these shifts on credit decisions and risk management can be significant:

- Changes in attribute distributions could affect the predictive power of your credit scoring models
- Shifts in population characteristics might necessitate model recalibration
- Errors in data interpretation could lead to inaccurate risk assessments and potential regulatory issues

Complementary Monitoring Techniques

In addition to Attribute Control Charts (ACCs), Population Stability Index (PSI), Characteristic Analysis Reports, and Scorecard Performance Reports each offer unique strengths and can be employed in tandem for comprehensive data monitoring.

Population Stability Index (PSI): PSI is essential for detecting shifts in the distribution of borrower attributes over time. It is particularly useful post-migration to identify significant changes in the borrower population that might impact model performance. PSI helps quantify these shifts, ensuring that the credit models remain accurate and reliable. PSI is ideal when you need to understand broad population changes and their potential impact on model performance.

Characteristic Analysis Reports: These reports provide a granular view of specific attribute changes, comparing their distributions before and after a change, such as a cloud migration. They are invaluable for pinpointing which attributes have shifted and understanding the nature of these shifts. This makes them an excellent choice for targeted investigations and corrective actions. Use Characteristic Analysis Reports when you need to drill down into individual attribute changes rather than broad population shifts.

Scorecard Performance Reports: These reports assess the overall effectiveness of credit scoring models by tracking metrics like the Gini coefficient and KS statistic. They measure the discriminatory power of models and are crucial for validating that models continue to perform as expected post-migration. These reports are best utilized when sufficient post-migration data is available to ensure that any changes in attribute values do not negatively impact the effectiveness of risk assessment tools.

By leveraging these complementary techniques, financial institutions can ensure robust monitoring and maintain the integrity of their credit decision-making processes during and after cloud migrations and other events that may cause shifts in borrower data.

Population Stability Index (PSI)

The Population Stability Index is a statistical tool used to detect shifts in the distribution of attributes over time. PSI is particularly useful for identifying when the population that your credit models is scoring has changed significantly from when the models were developed. One example that may cause such an impactful change is the transition to a cloud-based system.

Example: If the distribution of a key attribute, such as the number of 90+ days delinquencies, shows a marked change post-migration, a PSI calculation can help quantify this shift. A high PSI value would signal that the population has changed enough to potentially impact model performance, necessitating a closer examination and a possible recalibration of your scoring model.

Intention and Value: PSI helps ensure that your credit scoring models remain accurate by flagging significant changes in the borrower population, which might otherwise go unnoticed. This is especially critical immediately following a migration, as it can help identify whether data changes are due to the transition or underlying shifts in the population.

PSI Formula

Calculations and Data Requirements:

- **Data Required:** Distribution data of key attributes (e.g., delinquency counters, inquiry counts) both before and after migration.
- **Calculations:**
  1. **Divide the attribute's distribution** into bins or ranges (e.g., number of delinquencies: 0, 1, 2, 3+).
  2. **Calculate the proportion** of the population in each bin for both pre-migration and post-migration periods.
  3. **PSI Formula:**

$$PSI = \sum_{i=1}^{n} \left(\boldsymbol{P}_{pre}(i) - \boldsymbol{P}_{post}(i) \right) \times \ln\left(\frac{\boldsymbol{P}_{pre}(i)}{\boldsymbol{P}_{post}(i)} \right)$$

Where $\boldsymbol{P}_{pre}(i)$ and $\boldsymbol{P}_{post}(i)$ are the proportions of the population in the i^{th} bin for the pre- and post-migration periods, respectively.

  4. **Interpretation:** A PSI value less than 0.1 suggests no significant shift, 0.1 to 0.25 indicates a moderate shift, and values above 0.25 signal a significant change that could affect model performance.

Example of PSI Calculation

Imagine monitoring the number of 90+ day delinquent tradelines. The distribution before migration (pre) and after migration (post) might look like this:

Number of Delinquencies	Pre-Migration Proportion (P_{pre})	Post-Migration Proportion (P_{post})
0	0.70	0.65
1	0.20	0.25
2	0.07	0.08
3+	0.03	0.02

Calculating the PSI:

$$PSI = (0.70 - 0.65) \times \ln\left(\frac{0.70}{0.65}\right) + (0.20 - 0.25) \times \ln\left(\frac{0.20}{0.25}\right) + (0.07 - 0.08) \times \ln\left(\frac{0.07}{0.08}\right) + (0.03 - 0.02) \times \ln\left(\frac{0.03}{0.02}\right)$$

$$PSI = 0.05 \times 0.0741 + (-0.05) \times (-0.223) + (-0.01) \times (-0.134) + 0.01 \times 0.405$$

$$PSI = 0.00371 + 0.01116 + 0.00134 + 0.00405$$

$$PSI = 0.0203$$

This value indicates no significant shift (PSI < 0.1).

Characteristic Analysis Reports

Characteristic Analysis Reports compare the distributions of specific attributes before and after a change—in this case, the migration to a cloud-based system. This report is useful for pinpointing exactly which attributes have shifted and understanding the nature of those shifts.

Example of Inquiry Count Analysis

If inquiry counts are suspected to be off post-migration, a Characteristic Analysis Report can compare the distribution of these counts before and after the migration. This report will highlight whether the change is due to a reporting issue or a genuine shift in borrower behavior.

Intention and Value: These reports provide a granular view of how specific attributes are affected by the migration, allowing you to target investigations and corrective actions more effectively. It's a tool for deep-diving into individual attribute changes rather than broad population shifts.

Calculations and Data Requirements:

- **Data Required:** Attribute data distributions before and after the migration, typically in a binned format.
- **Calculations:**
  1. **Create Frequency Distributions:** Calculate the frequency (or percentage) of each attribute's value in the pre- and post-migration data.
  2. **Compare Distributions:** Visual or statistical comparison (e.g., chi-square test) to determine if there is a significant difference in the distributions.
  3. **Analysis:** Identify which specific attributes or bins show the most significant changes, helping to pinpoint potential issues or necessary adjustments.

Example of Characteristic Analysis Report

Imagine analyzing credit card utilization before and after migration. This report shows a shift in the distribution of credit card utilization, with more borrowers falling into the 0 - 30% range post-migration.

Credit Card Utilization (%)	Pre-Migration Proportion (%)	Post-Migration Proportion (%)
0-30	50	55
31-60	30	25
61-90	15	15
91-100	5	5

Scorecard Performance Reports

Scorecard Performance Reports evaluate the overall effectiveness of your credit scoring models by tracking metrics such as the Gini coefficient and the KS (Kolmogorov-Smirnov) statistic. These metrics assess the discriminatory power of your models, determining how well they differentiate between good and bad risks. Additionally, lift charts provide another valuable method for measuring scorecard performance, offering insights into the model's ability to identify positive and negative outcomes effectively.

Example of Scorecard Performance Reports

Example: After accumulating sufficient post-migration data, you might find that your scorecard's Gini coefficient has declined, indicating that the model is less effective at separating high-risk from low-risk borrowers. This could be due to changes in how data is reported in the cloud environment.

Intention and Value: While this report may require more time to implement, as it depends on the availability of sufficient post-migration performance data, it is essential for validating that your credit scoring models continue to perform as expected. It helps ensure that any changes in attribute values are not negatively impacting the overall effectiveness of your risk assessment tools.

Calculations and Data Requirements:

- **Data Required:** Performance data on borrower repayments (e.g., default vs. non-default) and corresponding credit scores over time, both pre- and post-migration.
- **Calculations:**

  1. **Gini Coefficient Calculation**
     - The Gini coefficient ranges from 0 to 1, with 0 indicating perfect equality (no discrimination) and 1 indicating perfect inequality (complete discrimination). In credit risk modeling, a higher Gini coefficient signifies better model performance in ranking borrowers by creditworthiness.
     - **Rank borrowers** by their credit scores.
     - **Calculate cumulative percentages** of good and bad loans.
     - **Plot these percentages** on a Lorenz curve and calculate the area between the curve and the line of equality.
     - **Gini Formula:**

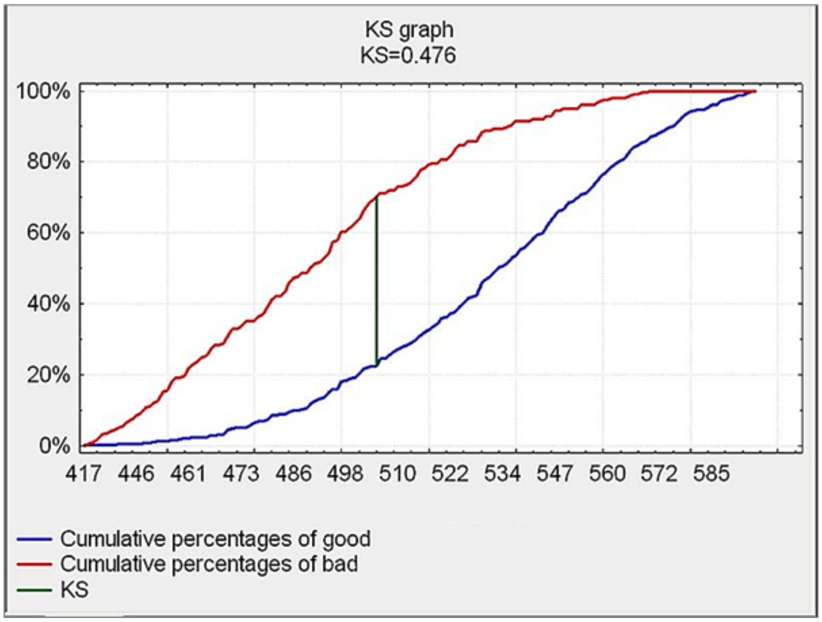
       $$\text{Gini} = 2 \times (\text{Area under the Lorenz curve}) - 1$$

  2. **KS Statistic Calculation**
     - **Divide the scored population** into deciles (or another appropriate segmentation).
     - **Calculate the cumulative percentage** of good and bad borrowers in each decile.
     - The KS Statistic is defined as the maximum vertical distance between the cumulative distribution curves of good and bad borrowers.
     - **KS Definition:**

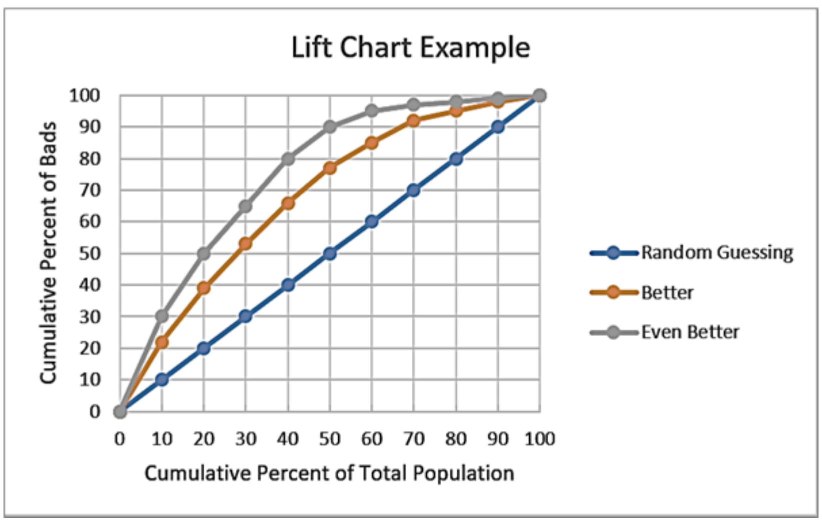
       $$\text{KS} = \max \left| \boldsymbol{F}_{\text{good}}(x) - \boldsymbol{F}_{\text{bad}}(x) \right|$$

     Where $\boldsymbol{F}_{\text{good}}(x)$ and $\boldsymbol{F}_{\text{bad}}(x)$ are the cumulative distribution functions of good and bad borrowers, respectively.

KS Graph



Lift Chart Example



- Lift charts are visual aids for measuring model performance.
- The x-axis shows the cumulative percent of the total population ranked from lowest to highest score by 10% increments (deciles).
- The y-axis shows the cumulative percent of bads scoring at or below each of the 10 deciles.
- The diagonal line represents the results of random guessing and is the baseline against which to evaluate lift.
- The greater the area between the lift curve and the baseline, the better the model is at separating good loans from bad loans.
- For 10% of the population, an "even better" model will capture 30% of bads compared to 10% of bads with random guessing.

Implementing a Comprehensive Monitoring Strategy

To effectively monitor changes in borrower data during the transition to cloud-based platforms, consider the following multi-step approach:

1. **Data Collection:**
    - Continuously collect data on relevant borrower attributes
    - Ensure consistency and accuracy in data formatting and updates
    - Set up automated data pipelines to streamline the collection process

2. **Chart Creation and Analysis:**
    - Implement Attribute Control Charts for key attributes
    - Regularly review these charts, ideally on a weekly or bi-weekly basis
    - Set up automated alerts for any points that fall outside control limits

3. **Population Stability Index Calculations:**
    - Calculate PSI for critical attributes monthly or quarterly
    - Maintain a dashboard of PSI values for easy tracking over time

4. **Characteristic Analysis:**
    - Conduct detailed Characteristic Analysis Reports quarterly
    - Focus on attributes that show significant changes in PSI or ACC analyses

5. **Scorecard Performance Monitoring:**
    - Generate Scorecard Performance Reports monthly or quarterly
    - Track Gini coefficients and KS statistics over time

6. **Investigation and Action:**
    - Establish a cross-functional team to investigate significant shifts
    - Develop action plans for addressing discrepancies or model performance issues
    - Collaborate with credit bureaus and internal teams to resolve data-related issues

7. **Staff Training:**
    - Provide regular training sessions on interpreting monitoring results
    - Ensure team members understand the implications of data shifts on credit decisions

8. **Regulatory Compliance:**
    - Maintain documentation of all monitoring activities and findings
    - Be prepared to demonstrate the robustness of your monitoring approach to regulators

Data Visualization Best Practices

Effective data visualization is crucial for quickly identifying and communicating data shifts. Consider the following best practices:

1. Use consistent color coding across all charts and reports
2. Implement interactive dashboards for easy exploration of data
3. Include trend lines and moving averages to highlight long-term patterns
4. Use heat maps for visualizing complex distributions or correlations
5. Provide clear annotations and explanations alongside visualizations

Regulatory Considerations

Financial institutions must be mindful of regulatory expectations regarding data integrity and model risk management. Key considerations include:

1. Documentation: Maintain detailed records of all monitoring activities, findings, and actions taken
2. Model Validation: Ensure that your monitoring approach is incorporated into your overall model validation framework
3. Data Governance: Implement robust data governance practices to ensure data quality and consistency
4. Reporting: Be prepared to report on data integrity and model performance to regulators upon request

Collaborating with the Credit Bureaus

Collaboration with the credit bureaus is crucial, especially when significant changes in borrower data are detected. Here are some strategies for effective collaboration:

1. Establish dedicated points of contact with each credit bureau for migration-related issues
2. Set up regular check-ins during the migration period to proactively address concerns
3. Develop a standardized process for reporting and escalating data discrepancies
4. Participate in industry working groups or forums focused on credit data quality

Case Study:
Navigating a Data Shift During Migration

To illustrate the application of these monitoring techniques, consider the following hypothetical scenario:

ABC Bank noticed a sudden increase in the number of applicants with no reported delinquencies after a major credit bureau migrated to a cloud platform. The bank's monitoring system flagged this change through multiple channels:

1. Attribute Control Charts showed a significant drop in the average number of delinquencies reported.
2. PSI calculations for the delinquency attribute yielded a value of 0.28, indicating a significant shift.
3. Characteristic Analysis Reports confirmed a 15% increase in applicants with zero delinquencies.
4. Scorecard Performance Reports showed a decline in the Gini coefficient from 0.72 to 0.65.

Investigation revealed that the credit bureau's new cloud system was not correctly capturing historical delinquency data for a subset of consumers. By quickly identifying and addressing this issue, ABC Bank was able to:

1. Adjust their credit decisioning process to account for the temporary data discrepancy
2. Work with the credit bureau to correct the data reporting issue
3. Recalibrate their credit scoring model once correct data was available
4. Avoid potential compliance issues by documenting the entire process

This proactive approach helped ABC Bank maintain the integrity of their credit decisions and minimize risk during the migration period.

Conclusion

The migration of major credit bureaus to cloud infrastructure marks a transformative event in data management, bringing both opportunities and challenges. By adopting a proactive, multi-faceted approach to monitoring—leveraging tools like Attribute Control Charts, Population Stability Index, Characteristic Analysis Reports, and Scorecard Performance Reports—financial institutions can navigate this transition smoothly.

This comprehensive monitoring strategy ensures that credit assessment practices remain accurate, effective, and resilient, even in the face of significant technological shifts. As the financial industry continues to evolve, the ability to quickly detect, understand, and respond to data changes will be a key differentiator for successful risk management.

By partnering with experienced data management firms like Digital Matrix Systems (DMS), financial institutions can leverage cutting-edge tools and expertise to implement robust monitoring systems. This collaboration not only enhances data integrity and model performance but also positions institutions to take full advantage of the enhanced capabilities offered by cloud-based credit data platforms.

As we look to the future, the importance of data quality and model performance monitoring will only grow. Financial institutions that invest in comprehensive monitoring strategies today will be well-positioned to adapt to future changes in the credit landscape, maintain regulatory compliance, and make more informed, data-driven decisions.