**Compliance Analysis**

*Ethan Ericson*

**Regulatory Requirements Mapping**

* Compliance Requirements:
  + HMDA - Home Mortgage Disclosure Act
    - The main purpose of the HMDA is to increase the transparency of home related loans to better track the ability of institutions to meet the real estate needs of their communities. It also serves to ensure compliance with fair-lending laws
    - To be compliant with HMDA, I must ensure that the data I use is accurate, complete, and timely. This compliance requirement mainly applies to attributes like applicant demographics, loan purpose, loan amount, application and decision date, and property location.
  + GLBA - Gramm-Leach-Bliley Act
    - The GLBA regulates how financial Non-Public Information on consumers is collected and stored. The main goal of this requirement is to prevent private information about an individual’s financial/ socioeconomic position from becoming publicly available.
    - To be compliant with GLBA, I need to take special care to protect the confidentiality of attributes like annual income, demographic info (ethnicity, race), and credit health (fico score/public count). Data minimization and role-based access control (RBAC) are important tools in safeguarding this info.
  + FCRA - Fair Credit Reporting Act
    - The FCRA regulates how consumer credit information is handled. The main purpose of the FCRA is to ensure that consumer credit data is used accurately and responsibly to prevent false rejections and discriminatory practices. The FCRA also requires the appropriate use of adverse actions in order to keep unstable borrowers from being approved for loans they are likely to default on (2008, looking at you).
    - Attributes like fico score, bankruptcy flag, tradelines delinquent, and public records count fall under this act’s protection. When using this data I must make sure that I have permissible purpose, which would only be granted for mortgage underwriting, account review/servicing for existing customers of a financial firm, fraud detection, and compliance reporting. For the sake of not breaking the law, I am going to say that this synthetic data was generated with the intent of being used for account review which I was granted specific permission to conduct.
* 6 V’s Mapping:
  + HMDA
    - Volume
      * Since HMDA requires reporting on a large amount of loan data, Volume is a primary concern with financial firms regulated under HMDA.
    - Velocity
      * Much of the loan data stored in financial institutions is streamed, but there are also batch reports made semi-annually or annually, so organizations must manage the velocity of incoming data in order to balance both types of flow. By proxy, I must make sure that I keep velocity in mind as I refresh the data I use for my reporting, as new data on loans may be streaming as I attempt to refresh, putting pressure on the pipeline.
    - Variety
      * Because the HMDA requires many key attributes to characterize a mortgage, there is a lot of variety in the data that I need to process. Property date is a spatial coordinate while loan purpose is description that could have been processed from an unstructured data source. Thankfully, in this analysis, most data have been compiled into structured csv format.
    - Veracity
      * This is perhaps the V most important to the HMDA. HMDA requires accuracy, completeness, and timeliness. The synthetic data I’m using would have hypothetically been audited under the HMDA before I received it, but that does not mean that I can become lax in ensuring that the veracity of the data is upheld. If I fail to refresh the data that feeds my reports, I will fail to provide timely or accurate insights, so I need to make sure that all reports that I release are up to date and clearly timestamped for clarity.
    - Valence
      * This V is strongly tied to Veracity under HMDA. The tables that provide mortgage information must have robust connections to ensure that data is represented accurately. I can ensure compliance under this V by joining based on primary key and foreign key relationships.
    - Value
      * HMDA compliance is all about generating value for regulators that want to observe and act on trends in home loans. This data can be used to spot discriminatory loan practices, changes in the housing market, or changes in the economy. I can ensure that value is generated from this data by complying with the HMDA principles of accuracy, completeness, and timeliness.
  + GLBA
    - Volume
      * There is a lot of private information that needs to be protected under GLBA, so managing volume effectively while protecting consumer privacy is integral in my analysis.
    - Velocity
      * Upstream from me, the “collectors” of this synthetic data had to ensure RBAC was in place to manage the ebb and flow of requests to the data. The security of NPI has to be upheld in the face of potentially overwhelming tidal waves of requests, which could come at a moment’s notice if people have a standardized daily refresh time.
    - Variety
      * Some of the data in this dataset is not NPI, and therefore under GLBA does not require as careful oversight. As a data user, I need to make sure to account for the variety of non-NPI and NPI data, protecting sensitive information while still generating valuable insights.
    - Veracity
      * Data must be correct to prevent unintentional sharing of sensitive information. If a column is labeled incorrectly, a data user could mistake that attribute for non-NPI data and leave it unmasked.
    - Valence
      * Under GLBA, joins should be carefully created in order to avoid redefining user attributes or exposing NPI. Once again, PK’s and FK’s are the key.
    - Value
      * Following GLBA generates value for the users of the reports without exposing private customer information. Protecting NPI ensures continued trust from customers.
  + FCRA
    - Volume
      * With many tradelines per customer, and many customers per organization, credit data is massive and must be managed carefully to ensure fair reporting.
    - Velocity
      * With many transactions occurring daily per tradeline, the velocity of credit data is quite volatile. FCRA requires that despite this velocity, relatively fresh data should be used at key decision points. To comply with FCRA, I should refresh my data sources often.
    - Variety
      * There is some variety in credit data. Fico is a numeric measure, while bankruptcy flag is a boolean, but most of credit data is fairly structured.
    - Veracity
      * Under FCRA, accuracy of credit information is very important, and consumers have the right to dispute inaccurate claims. The potential of disputes requires that veracity be kept in mind at all times when handling credit data, as you need to back up your decisions and insights if they come under scrutiny.
    - Valence
      * Data from credit bureaus must only be connected to other financial related data when there is permissible purpose, which luckily I’ve granted myself!
    - Value
      * The appropriate use of credit data under FCRA ensures value for both borrowers and lenders. By protecting the accuracy of credit data, lenders avoid making risky loans, and borrowers have their credit history presented in a fair unbiased light to lenders. As I’m analyzing customer accounts, I will be sure to uphold the accuracy of the credit data to protect the value of my insights to lenders and borrowers.

**Risk Assessment Matrix**

* Risk Matrix:

*(I used Agent in Chat GPT to generate in excel)*

|  |  |  |
| --- | --- | --- |
| **Risk** | **Likelihood** | **Business Impact** |
| 1. Incorrect or incomplete HMDA reporting | Medium | High |
| 2. Unauthorized access to Restricted data (GLBA) | Low–Medium | High |
| 3. Failure to send accurate/timely Adverse Action Notices (FCRA) | Medium | High |
| 4. Inaccurate credit bureau data usage (FCRA) | Medium | Medium |
| 5. Data breach or unencrypted storage (GLBA) | Low | Very High |

* + Explanations:
    - If loan application data (income, demographic attributes, property info) is misreported, regulators may impose penalties and the institution’s reputation for fair lending could suffer.
    - If demographic or bureau data is accessed without authorization, it would violate privacy laws, trigger regulatory fines, and erode consumer trust.
    - If denied/incomplete applications don’t receive correct notices within required timelines, the lender faces regulatory sanctions, lawsuits, and reputational harm.
    - Using outdated or incorrect bureau info could lead to unfair denials or pricing, and exposes the institution to consumer disputes and legal actions.
    - A major breach involving NPI would cause heavy fines, lawsuits, and severe reputational damage.
* Critical Governance Areas & Mitigation
  + Access Control and Data Protection
    - GLBA places a large emphasis on protecting NPI. As I have just received this data, I have not set up any RBAC or data minimization procedures. It is extremely critical that I implement these measures before I make the data public in any form in order to comply with the GLBA.
  + Data Quality & Integrity
    - HMDA places a behemoth emphasis on data quality and integrity. Right now, I have not checked that data is consistent across the tables in my dataset, nor have I investigated if data is timely (I’m assuming I’m operating in 2024). I should do EDA on the data to check that it meets HMDA’s requirements for accuracy, completeness, and timeliness.
  + Consumer Rights Compliance
    - FCRA requires that consumers’ rights to an accurate and fair evaluation of their credit data be maintained. In order to ensure that this is respected, I should first validate that every loan decision that is rejected maps to an appropriate adverse action entry, then set up an automated workflow to use the data contract to make sure that every decision of reject maps to an adverse action entry.