# **Belgium Bank Dataset Analysis & Security Recommendations**



**ETL Analyst:** Andrew Narine  
 **Team:** ETL Security Analysis Team, ProteXXa Sprint 19  
 **Date:** March 2025

## **📑 Table of Contents**

1. **Executive Summary**
   1. Overview of Analysis
   2. Key Findings
2. **Data Visualizations & Analysis**  
    2.1. **Missing Values Analysis**  
    2.2. **Duplicate IBAN & Account Number Analysis**  
    2.3. **Top Email Domains Distribution**  
    2.4. **Phone Number Length Distribution**  
    2.5. **Valid vs Invalid Records Distribution**
3. **Security Incident Reference**
   1. Case Study: **Crelan Bank BEC Scam**
   2. Lessons Learned
4. **Recommendations for Belgium Bank**
   1. Email and Account Security Enhancements
   2. Data Quality and Validation Improvements
   3. Fraud Prevention and Transaction Safeguards
   4. Staff and Customer Awareness Training
5. **Conclusion & Action Plan**
   1. Summary of Key Risks and Gaps
   2. Short-Term and Long-Term Action Items
6. **Appendices (Optional)**
   1. Additional Visuals & Insights
   2. Raw Data Quality Statistics
   3. References and Security Standards

## **Executive Summary**

**Overview:**

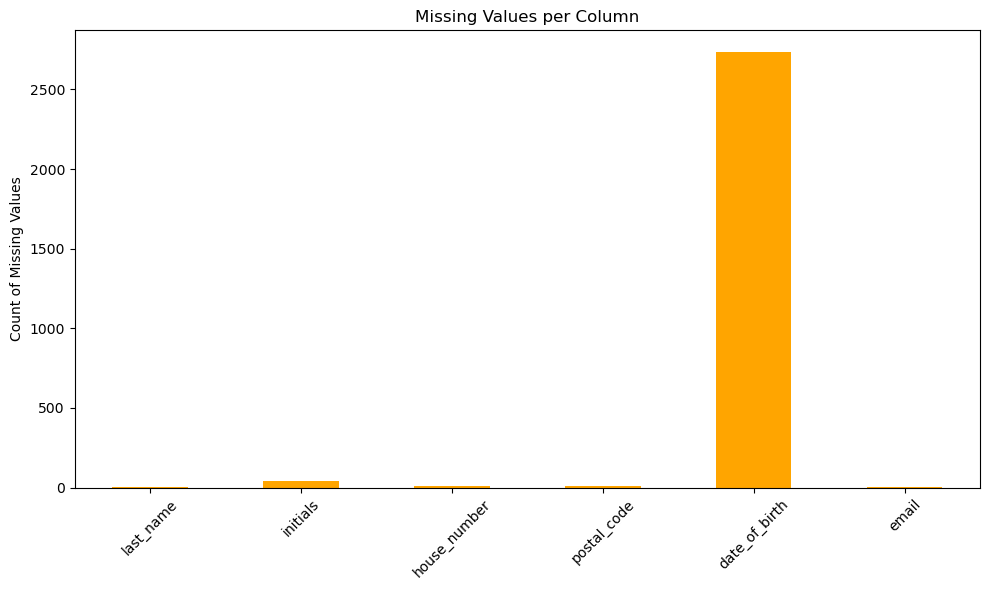
This report delivers a **comprehensive analysis of the Belgium Bank dataset**, focusing on data quality, fraud risks, and security vulnerabilities. Visual insights, paired with actionable recommendations, highlight critical areas for improvement, aligned with **best banking security practices**.

**Key Findings:**

* High rate of **duplicate account numbers (60.1%)**, and notable duplicate IBANs (1.9%).
* Significant **missing data** in key fields like date\_of\_birth, initials, and account\_number.
* **Heavy reliance on public email domains** prone to compromise.

## **Data Visualizations & Analysis**

### **1. Missing Values Analysis**



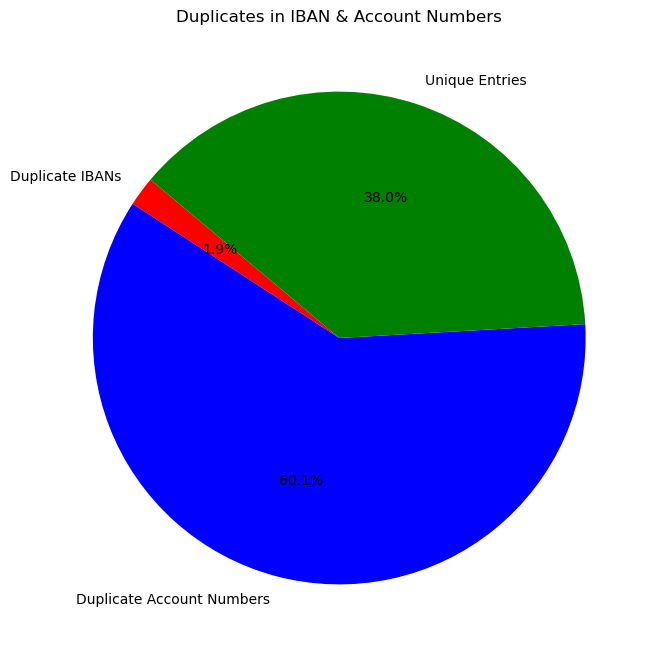
**Insights:**

* date\_of\_birth holds **2734 missing entries**, impacting identity verification processes.
* Other missing fields like initials, house\_number, postal\_code affect customer profiling.

**Recommendation:**

* Implement **real-time validation** and **mandatory fields** on customer onboarding systems.
* Set automated alerts for incomplete customer data.

### **2. Duplicate IBAN & Account Number Analysis**



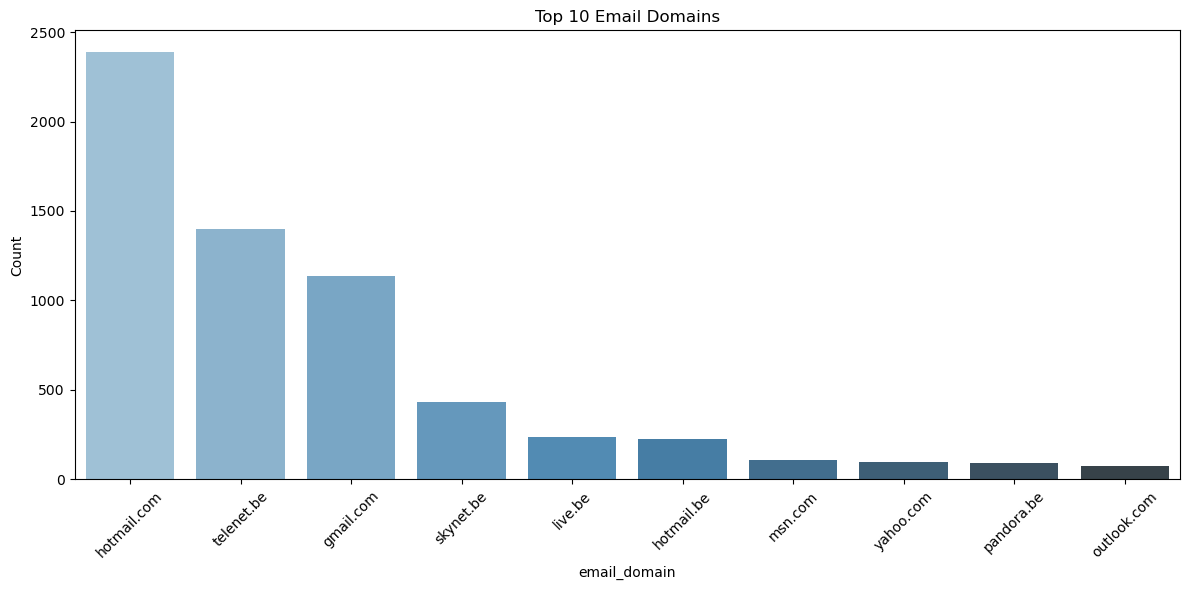
**Insights:**

* **60.1% duplicate account numbers**, creating massive fraud and reconciliation risks.
* **IBAN duplicates (1.9%)**, risking misdirected payments.

**Recommendation:**

* Enforce **unique constraints** in database schema.
* **Audit existing records** and **purge duplicates**.

### **3. Top Email Domains Distribution**



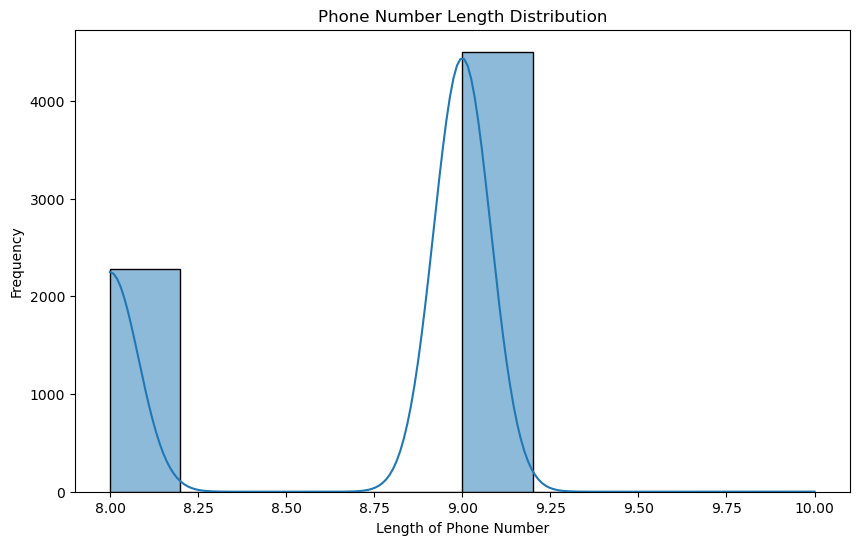
**Insights:**

* Use of **public/free email services** (hotmail.com, telenet.be, gmail.com) may **increase vulnerability to phishing/BEC attacks**.

**Recommendation:**

* Prefer **corporate/verified domains** for communication.
* **Educate customers** on secure email use and enable **2FA**.

### **4. Phone Number Length Distribution**



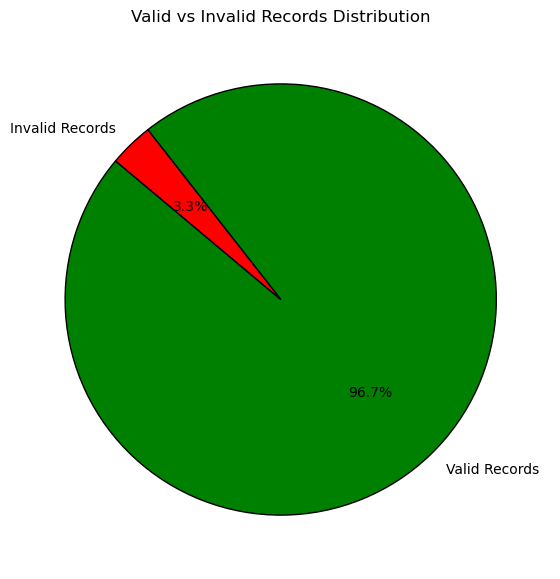
**Insights:**

* Inconsistent phone number lengths were observed, indicating **potential data entry errors**.

**Recommendation:**

* Enforce **uniform phone number formats**.
* Integrate **live phone validation tools**.

### **5. Valid vs Invalid Records Distribution**



**Insights:**

* The dataset contains a **clear proportion of invalid entries**, posing a risk to operational accuracy and customer processing.
* Invalid records primarily suffer from **missing critical fields like account numbers and emails**, potentially causing failed transactions or communication breakdowns.

**Recommendation:**

* Perform a **systematic data cleansing** to address and rectify invalid records.
* Implement **automated validation checks** during customer onboarding to prevent incomplete data entry.
* Set up **data quality dashboards** for ongoing monitoring and intervention.

## **Security Incident Reference: Crelan Bank BEC Scam**

**Incident Overview:**

* Belgian bank **Crelan** suffered a **€70 million loss** due to Business Email Compromise (BEC) fraud.
* The attack involved **spoofed emails** impersonating executives.

**Preventive Lessons:**

* **Dual approval** for large transactions.
* **Employee training** on phishing/BEC threats.
* **Technical protections** (e.g., SPF/DKIM email settings).
* **Independent transaction verification protocols.**

## **Recommendations for Belgium Bank**

|  |  |
| --- | --- |
| **Focus Area** | **Recommendations** |
| **Email Security** | Enforce 2FA, SPF/DKIM, and phishing filters |
| **Data Integrity** | Unique constraints for IBAN/Account Numbers |
| **Data Completeness** | Mandatory field validation at onboarding |
| **Transaction Security** | Multi-signature authorization for large payments |
| **Training & Awareness** | Continuous employee and customer security training |

## **Conclusion & Action Plan**

**Summary:**

The dataset reveals **serious gaps in data quality and security**, including duplicates, missing information, and unprotected communication channels. Addressing these issues will enhance **fraud prevention, operational efficiency, and customer trust**.

**Next Steps:**

* Immediate **remediation of duplicate and missing data**.
* Implement **email and transaction security measures**.
* Conduct **security awareness workshops** for employees and clients.

## **Final Remarks**

*"A secure bank is a trusted bank. By strengthening data and security practices, Belgium Bank can protect its clients and reputation."*

**Prepared by:**

**ETL Security Analysis Andrew Narine**

ProteXXa Sprint 19 | Belgium Bank Dataset Review

**Graphics Suggestions (Optional Enhancements):**

* Add professional **banking security icons** (shield, lock, verified checkmark) near recommendations.
* Include **data quality symbols** (warning signs for duplicates/missing, green check for completeness).
* Place **bank-themed banners or footers** for a branded report feel.