

# KYC & Client Onboarding Data Quality Audit

*Amal S*



# Agenda



Introduction



Primary Goals



Areas of  
Growth



Timeline



Summary

# Introduction

Know Your Customer(KYC) compliance is a core requirement in banking and financial services. Ensuring accurate and complete client data helps institutions prevent fraud, meet regulatory obligations, and improve operational efficiency.

This project simulates a real-world scenario of client onboarding, where incomplete or inconsistent data can trigger compliance risks.

Using Python, I created synthetic client records and implemented logic to detect missing values, data entry errors, and sanctioned country flags.

The project also includes fuzzy matching techniques to correct misspelled country names — a common issue in data entry.

Visual summaries of flagged issues provide actionable insights for improving KYC data quality and onboarding workflows.

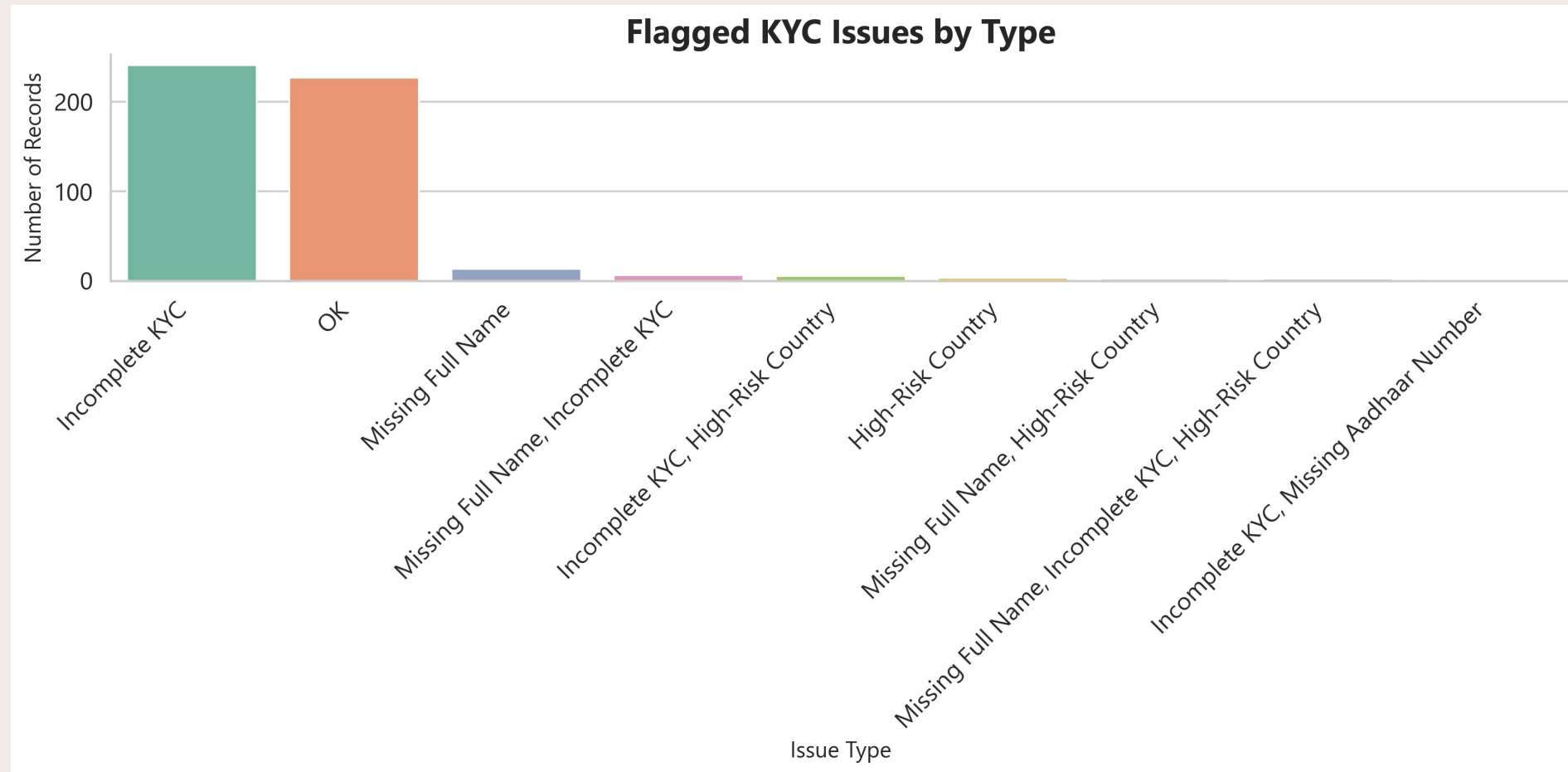


# Primary goals

*Ensure completeness, accuracy, and compliance of client onboarding data using Python-based validation and flagging logic.*



# Flagged KYC Issues by Type



# KYC Data Risk Overview

| Issue Type   | Number of Records | Risk Level | Suggested Action                                  |
|--|-------------------|------------|---|
| Incomplete KYC                                       | 241               | ● High     | Improve follow-up and automate document reminders |
| Missing Full Name                                    | 13                | ● Moderate | Make full name mandatory at entry point           |
| Missing Full Name, Incomplete KYC                    | 6                 | ● High     | Flag for manual review before account activation  |
| Incomplete KYC, High-Risk Country                    | 5                 | ● High     | Escalate for compliance officer review            |
| High-Risk Country                                    | 3                 | ● High     | Block or require enhanced due diligence (EDD)     |
| Missing Full Name, High-Risk Country                 | 2                 | ● Medium   | Cross-check with source documents                 |
| Missing Full Name, Incomplete KYC, High-Risk Country | 2                 | ● Critical | Multi-risk: escalate immediately                  |
| Incomplete KYC, Missing                              | 1                 | ● High     | Indian client: Aadhaar                            |

“

Without data, you're just  
another person with an  
opinion

*W. Edwards  
Deming*

”

# Tools & Technologies Used

| Tool/Tech          | Purpose                          |
|--------------------|----------------------------------|
| Python (pandas)    | Data cleaning, simulation        |
| Faker              | Generate synthetic client data   |
| RapidFuzz          | Correct misspelled country names |
| Matplotlib/seaborn | Visualizations                   |
| PowerPoint         | Final reporting & presentation   |



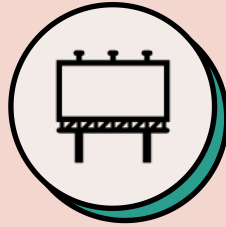
# Key Takeaways / Lessons Learned

- Real-world KYC processes can have **hidden data quality issues**
  - Simple Python tools can help detect **compliance risks at scale**
  - Fuzzy matching and rule logic are powerful for **automated anomaly detection**
  - This project strengthened my confidence in **data-driven decision-making**
-

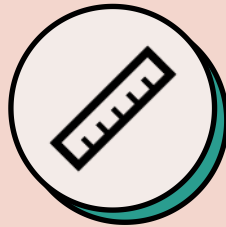
# Strategic Roadmap for KYC Data Integrity



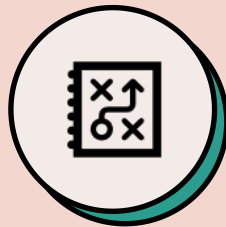
Planning  
*Define Key Fields & Flags*



Validation  
*Build Validation & Simulation*



Cleaning  
*Data Correction & Enrichment*



Analysis  
*Detect Patterns & Risks*



Reporting  
*Deliver Insightful Outcomes*

# Timeline

Sep 20XX

*Synergize  
scalable  
e-commerce*

Jan 20XX

*Coordinate e-  
business  
applications*

May 20XX

*Deploy  
strategy  
networks with  
compelling e-  
business needs*

Nov 20XX

*Disseminate  
standardized  
metrics*

Mar 20XX

*Foster  
holistically  
superior  
methodologies*

# Areas of focus

## B2B market scenarios

- Develop winning strategies to keep ahead of the competition*
- Capitalize on low-hanging fruit to identify a ballpark value*
- Visualize customer directed convergence*

## Cloud-based opportunities

- Iterative approaches to corporate strategy*
- Establish a management framework from the inside*

# How we get there

## ROI

*Envision multimedia-based expertise and cross-media growth strategies*

*Visualize quality intellectual capital*

*Engage worldwide methodologies with web-enabled technologies*

## Niche markets

*Pursue scalable customer service through sustainable strategies*

*Engage top-line web services with cutting-edge deliverables*

## Supply chains

*Cultivate one-to-one customer service with robust ideas*

*Maximize timely deliverables for real-time schemas*



# Summary

*At Contoso, we believe in giving 110%. By using our next-generation data architecture, we help organizations virtually manage agile workflows. We thrive because of our market knowledge and great team behind our product. As our CEO says, "Efficiencies will come from proactively transforming how we do business."*

# Thank you

*Mirjam Nilsson*

*mirjam@contoso.com*

*www.contoso.com*

