

Project Introduction

Data security is crucial in the digital age, marked by the adoption of cloud solutions [1]. As organisations in the UK shift to cloud solutions [2] for convenience, scalability, and flexibility, they also face new vulnerabilities. The increasing use of cloud platforms for storing personally identifiable information (PII) demands strong security measures. Cloud data loss prevention (DLP) can detect and prevent the exposure of sensitive data like PII. A government project is now assessing DLP's efficiency in detecting and managing sensitive text using both standard and custom rules.

Project Aims

- To research and configure a cloud based DLP system to scan and detect specific sensitive data.
- To determine how successful DLP implementation is at identifying and removing sensitive data.
- To devise methods for circumventing specific DLP control and construct recommendations against this.

Project Objective

- To showcase the specific function of DLP to protect and remove sensitive data.

Custom Info-types

The following syntax patterns are used to detect the PII in the test datasets when performing an inspection job using the custom info-types.

National_Insurance_Custom_Syntax

(\W^)([A-CEGHJ-NPR-TW-Za-ceghj-npr-tw-z][A-CEGHJ-NPRT-WX-Za-ceghj-nprt-wx-z])[-\s]{0,1}\d{2}[-\s]{0,1}\d{2}[-\s]{0,1}\d{2}[-\s]{0,1}[A-Da-d]

UK_driving_custom_syntax

[A-Za-z]{5}[-\s]{0,1}\d{6}[-\s]{0,1}[A-Za-z]{2}[-\s]{0,1}\d{1}[-\s]{0,1}[A-Za-z]{2}

UK_Passport_Syntax

\d(?:\s?\d){8}

Test Dataset Formats

Merged:

- TAYLO543210VH5GD, taylo543210vh5gd, 263856289, BS521956A and bs521956a.

Spaces:

- TAYLO 543210 VH 5GD, TAYLO 543210 VH5GD, taylo 543210 vh 5gd, taylo 543210 vh5gd, bs 52 19 56 a, BS 52 19 56 A and 2 6 3 8 5 6 2 8 9.

Dashes:

- taylo-543210-vh5gd, taylo-543210-vh-5gd, TAYLO-543210-VH5GD and TAYLO-543210-VH-5GD and BS-52-19-56-A, bs-52-19-56-a.

Implementation

- Create a Google Cloud Platform (GCP) project and enable the Cloud DLP API.
- Upload the test dataset to a Cloud Storage bucket.
- Create a test folder in another storage bucket.
- Create custom info-types to detect PII data: National insurance, driving license and passport numbers.
- Create an inspection job to scan the test dataset for sensitive PII data using custom or stored info-types.
- Review the results of the Cloud DLP job to determine how much sensitive data was detected and if it was remediated.

Examples of Successful Remediation

National Insurance Number: PT-23-45-67-F

263856289 or 2 6 3 8 5 6 2 8 9

driving license MORGA657054SM9IJ will be revoked

Customer Letter - MORGA-657054-SM-9IJ

➡

➡

➡

➡

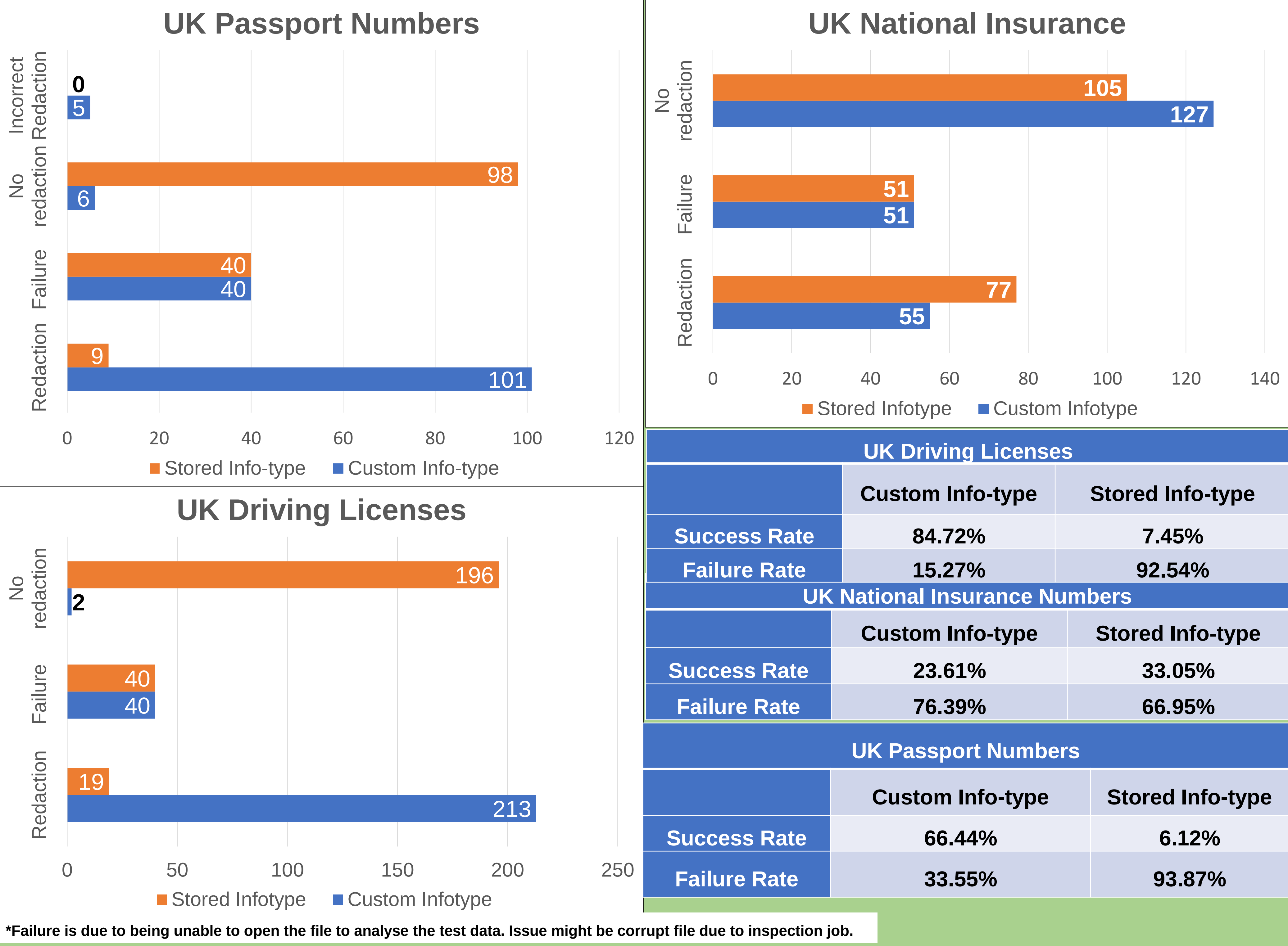
National Insurance Number: <SENSITIVE_DATA_REMOVED>

<SENSITIVE_DATA_REMOVED> or <SENSITIVE_DATA_REMOVED>

driving license <SENSITIVE_DATA_REMOVED> will be revoked

Customer Letter - <SENSITIVE_DATA_REMOVED>

Project Test Results



Conclusion

The project was successful in achieving its objective and aims. It researched and configured a cloud-based DLP system to scan and detect specific sensitive data, determined how successful DLP implementation is at identifying and removing sensitive data and tested methods for circumventing specific DLP controls.

Custom info-types outperformed stored info-types in identifying passport numbers and driving licenses. However, the stored info-type exceeded the custom one for national insurance numbers. Google must adjust the DLP's driving license format to serve its UK customers better. The poor performance of the custom info-type for national insurance numbers may be down to errors in the syntax. Throughout the evaluation, 26 inspection jobs were executed for each PII info-type, amounting to 156 in total. The consistency of results was noteworthy, with no variations observed. However, it was determined that file format corruption, especially in HTM and XML, particularly impacted the outcomes.

Further Thoughts

- Assign more time to the project for a more extensive review and customisation of the custom info-type rules.
- Expand the test to include a broader range of the DLP's supported file formats.
- Explore the detection and remediation of images.
- Explore issues encountered by file corruption in the scanned data file.
- Explore other forms of PII that the DLP system could protect.

[1] M. K. Pratt, "The importance of data security in the enterprise," August 2022. [Online]. Available: <https://www.techtarget.com/searchsecurity/feature/The-importance-of-data-security-in-the-enterprise>.
[2] T. O. o. C. (Ofcom), "Cloud services market study - Interim report," The Office of Communications, 2023.