Issues in Information Systems
Volume 23, Issue 2, pp. 242-254, 2022

DOI: https://doi.org/10.48009/2_iis_2022_120

Serverless data pipeline for data-driven decision making to mitigate pet surrenders

Adedamola Ismail Sanusi, *University of St. Thomas, ola.sanusi@stthomas.edu*
Subhransu Sekhar Nanda, *University of St. Thomas, nanda.s@stthomas.edu*
Sumit Kumar Panjwani, *University of St. Thomas, sumit.panjwani@stthomas.edu*
Safwan Islam, *phData Inc., safwan@phdata.io*
Manjeet Rege, *University of St. Thomas, rege@stthomas.edu*

Abstract

Secondhand Hounds (SHH) is a non-profit animal rescue organization located in Minnetonka, Minnesota and was facing a rise in pet surrenders. A pet surrender is when an owner relinquishes ownership of the pet to a shelter or rescue. On top of this, the processing of surrender applications was a manual and extremely tedious process. The aim of this project was two-fold: create an automated data pipeline and a reporting dashboard for key insights. This would enable data-driven decision making for SHH to hopefully mitigate these surrenders in the future. As part of this project, we collaborated with the team at Secondhand Hounds to automate the processing of surrender applications and storing of records, created a centralized NoSQL database for surrender data and built a business intelligence dashboard for data analysis.

Keywords: Non-profit, Pet Surrenders, Cloud Computing, ETL Automation, Data Validation, Data-Driven Decision Making

Problem Description

The project was a data for good initiative for Secondhand Hounds. Secondhand Hounds provides services for at-risk surrendered animals received from the public. They provide veterinary care and daily necessities for rescued/returned dogs, cats and other animals that deserve another chance at finding a loving and safe home. Their rescue services also include on-site sheltering and finding of foster homes for rescued/returned pets. Their end goal is finding surrendered pets permanent, loving adoption homes.

Secondhand Hounds' work processes, and current information system brought up challenges in accessing and utilizing their data effectively. They had a limited approach for keeping track of information, generating reports and no way of performing analytics on historical records. For example, parts of their workflow for capturing, processing, and storing relevant data were being conducted manually. Secondhand Hounds received data through filled out pet surrender applications on its website. Secondhand Hounds' website was created on a WIX platform, a cloud-based web development service provider. After applications are submitted, auto generated emails send a combination of surrender application data, a comma-separated-values (CSV) file and picture attachments to an organizational email. This data emailing system was used to keep a log of application events. The surrender administrator was the role whose job duties included processing all surrender applications and coordinating with foster-care coordinators as shown in Figure 1. Extracting data from surrender applications was a manual process. Data was manually copied from the WIX website to Google sheets for processing by the surrender administrator. Processing a surrender application was mostly concluded after a decision was made on that application. That application data then becomes a

historical record. Historical records were being stored on Excel spreadsheets and on Google sheets. With this process there was a high likelihood of human errors and unintended changes to data while copying, processing current data, and saving historical data on Google sheets or Excel files. It was difficult to analyze or visualize historical data with this approach. Also, this difficulty in analyzing and visualizing data was further compounded by the way data was structured and by the format of field entries. These issues and limitations prevented them from identifying any patterns or trends in their data. Added to these complications was the fact that they had no in-house IT team. They outsourced their IT needs to a third-party technology services company on a contract-based ticketing system.

Pet surrenders can be a devastating experience for pets and owners but also can be grueling for pet rescue organizations. This is related to their efforts in keeping pets in shelters temporarily, relocating them to foster homes or getting them adopted into permanent homes. This experience is exacerbated when there is a spike in pet surrenders. According to Glass and Palmer (2021), animal shelters have recently seen a surge in pet surrenders across the US. This ongoing hypothesis according to Glass and Palmer (2021) is that this is due to the pandemic. Glass and Palmer (2021) went on to state that many Americans had adopted a record number of pets in 2020 for emotional support or companionship as most of the people had to work from home; most of the businesses, offices and schools are reopening, and these pet owners are finding less time to take care of their pets resulting in the spike in surrenders (Glass & Palmer, 2021).

The primary challenge for Secondhand Hounds was to address an ongoing spike in the number of pet surrender applications. To do this, they needed to understand what was happening, why this was happening and then to plan and execute mitigation strategies. To address these key issues, the data from surrender applications would need to be collected and analyzed.

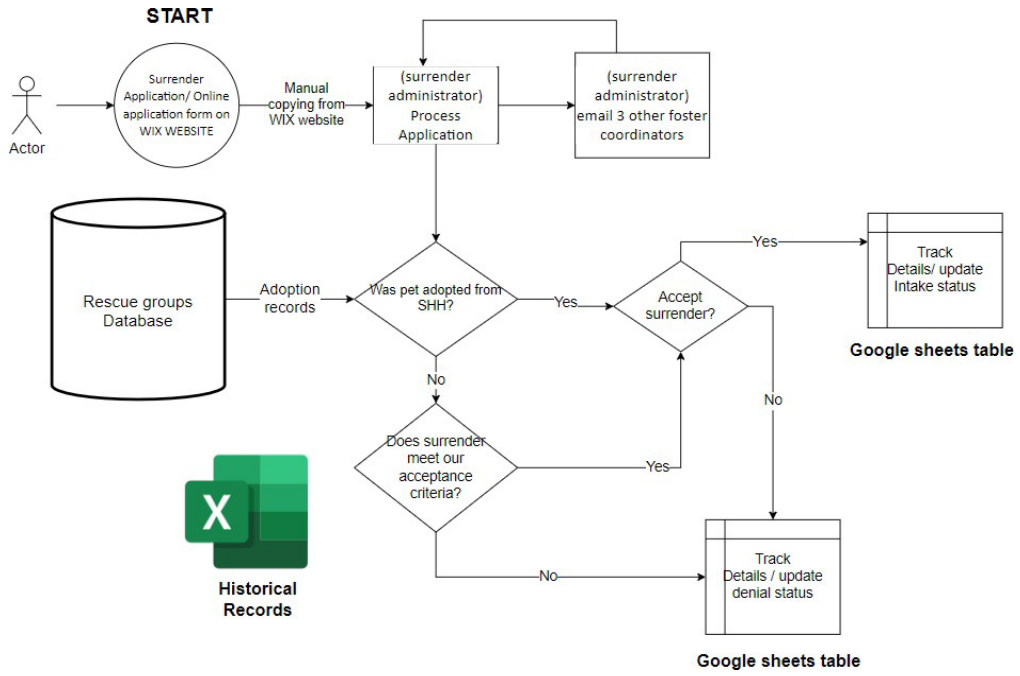

Figure 1. Current process flow diagram

Literature Review

Ho et al. (2021) presented a study to determine if there has been a significant increase in the pet adoption during the initial period of the pandemic. Relative search volume (RSV) was used to determine the score from 0(lowest) to 100(highest). The countries which contributed to the data are Australia, the United States, Canada, New Zealand, the United Kingdom, Singapore, the Philippines, and Malaysia. They found that there was a multi-fold increase in pet, dog, and cat adoption between April to May 2020, in the early pandemic phase compared to the same period in 2019. Also, there was a concern regarding the possible abandonment of newly adopted pets once the post pandemic restrictions were removed.

Dinata et al. (2021) solved a similar problem as this project. They improved the manual process used by The Indonesian Red Cross - Palang Merah Indonesia (PMI) to collect data from blood donors. According to Dinata et al. (2021), these data which have been manually collected were later entered into a relational database. They created a website for data collection from donors to replace the manual process. They built a dashboard to help an admin see various data points related to different data sources and donor data. They adopted a Waterfall approach which included different testing methodologies to test the effectiveness of their implementation.

Eid et al. (2018) identified four criteria for choosing a cloud service delivery model. They identified these as: available budget, flexibility for changing user and organizational requirements, IT skills available to an organization and the urgency of the need.

Koparkar and MacKrell (2015) focused on how not-for-profits (NFPs) can leverage business intelligence (BI) in the cloud while navigating the risks and vulnerabilities related to cloud governance, the access to and security of organizational data. They went on to state that as important as the benefits of cloud computing are for NFPs, issues such as proper governance and security controls need to be addressed together with the technological needs for moving to the cloud. Thus, using a trusted cloud platform is a necessity (Koparkar & MacKrell, 2015).

Maxwell et al. (2016) in their study on data-driven decision making (DDDM) by non-profits defined DDDM activities as the collection, analysis and applying of data to decisions. They identified two dimensions for data-driven decision making: the culture at the organization and the perceptions of the staff. Kline and Dolamore (2020) in their case study on the data-driven culture at a mission-driven organization, found that the value of data will only be realized when it gets used. They concluded that active management of data which covers its collection and analysis, improved perspectives on the importance of data for organizational decision making.

Objectives

Improved Process: Requirements for the proposed system were gathered using questionnaires, interviews via Zoom meetings and from analyzing the documents that the Secondhand Hounds team provided. From the requirements gathered, features and requirements of the improved process were identified as follows:

1. Automated extraction, transformation, and loading (ETL) of surrender application data submitted on Secondhand Hounds' website to a centralized database.
2. Data cleanup and loading of historical data to a centralized database.
3. Data validation or integrity checks and errors identification during the ETL process.
4. Centralized storage using a cloud platform.
5. Ability for a user to add data manually to update existing applications with little technical support. This would be in the event of a failure of the automated system which populates data from Secondhand Hounds' website.

6. Search capability to find any given surrender application from the database.
7. Business Intelligence reports as shown in table1, that will help Secondhand Hounds understand various trends and factors affecting surrenders.

Methodology

Selection of cloud service provider

As part of the technical solution, it was important to understand Secondhand Hounds' needs in order to choose an appropriate cloud service provider. An assessment was done to compare various offerings of two major cloud service providers, identified as Microsoft Azure and Amazon Web Services in line with Koparkar and MacKrell (2015).

The factors that were identified as being key to cloud adoption by the Secondhand Hounds team were majorly related to budget and tech support (Eid et.al, 2018). Applicable in this regard included the free tiers being offered by the cloud service providers, non-profit grants, levels of tech support offerings, and the applicability of non-profit grants towards tech support offerings. These assessments added to the fact that Secondhand Hounds had an existing non-profit account with Microsoft, helped them decide on a cloud provider. The decision was made to go forward with Microsoft Azure cloud platform.

Data cleaning, restructuring, and loading of historical data to Cosmos DB

Document-oriented model restructuring: Multiple Excel tabs were consolidated to restructure the surrenders spreadsheet. The consolidation was done in line with a denormalized model for the NoSQL database implementation.

Data loading and cleaning: Historical records were cleaned and transformed using Python Pandas library. Field transformations were carried out as follows:

- ***Field Mapping transformations:*** Mapping original values to their transformed format.
- ***Date field transformations:*** Python's regular expressions, re module and datetime module were used to check the submission date field for valid date formats and flag non-conforming entries.
- ***Splitting of entries:*** This involved splitting of applicants' names and multiple combined surrender applications.
- ***Categorical fields and numeric transformations:*** This involved transforming abbreviations into their required values, and parsing numbers from text entries. Parsed numbers were used to generate new columns with a floating-point data type for "Age" and "Weight" entries. A categorical field was created to transform "Reason for Surrender" entries into categories that can be measured. Other categorical fields were created, and keywords were parsed to transform the original text entries into their categorical equivalent.

Table 1. List of reporting requirements

Reporting requirements	
Number of surrender applications per day/month/year	Average age of surrenders
Number of surrenders by species per day/month/year (dogs vs cats)	Number of surrender applications denied per day/month/year
Number of surrenders by size per day/month/year	Top reasons for denial per month/year
Top reasons for surrender by category per month/year	Length of time between application date and intake date
Number of surrenders by age range of pet	Number of Owner-surrender vs Adoption Returns per month/year

Data Extraction and Validation processes

Data extraction: The extraction, transformation, and loading (ETL) of data to Cosmos DB occurs in two phases: Phase 1 for extraction, transformation and loading and Phase 2, for validation, extraction and updating. Microsoft's Azure Functions service was used to carry out both phase 1 and phase 2 implementations. Azure Functions is an on-demand serverless compute which can be used to respond to events and carry out automated workflows (Microsoft, 2022a). In a serverless approach, the need for infrastructure provisioning, scaling and management is handled by a cloud service provider (Microsoft, 2022b). This brings advantages of increased productivity, faster delivery of solutions and increased focus on delivering core values to an organization for developers (Microsoft, 2022b). Azure Functions services were used to respond to events and carry out processes as shown in figure 1, figure 2 and figure 5. Automated deployments of the blocks of Python code for these processes were carried out using Azure Functions' Consumption plan. The consumption plan is a pricing plan that is charged for only when codes are run (Microsoft, 2022c). This was the most suitable option for this project scenario due to its pay-as-you-go pricing model and a monthly free grant which allowed for a cost-effective solution.

Phase 1 Extraction, Transformation and Loading

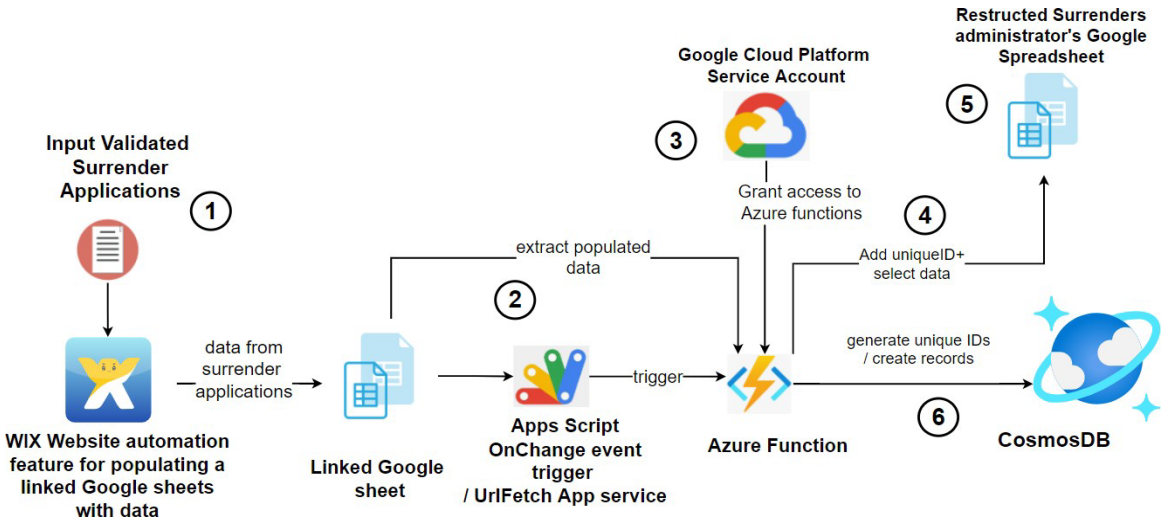

Figure 2. Phase 1 extraction, transformation, and loading

1. An automation feature on Secondhand Hounds' website enables data to be populated to a linked Google sheet whenever an application for surrender is submitted.
2. An Azure function is triggered when data is added to rows on the sheet. Google Apps Script was used to extend the Google sheets' functionality with automations. With Apps Script, a trigger can be used to connect third-party services external to Google products (Google, 2021). The triggered Azure function accesses data on the Google sheets via Python's gspread library.
3. Authentication for connecting an Azure Function to Google sheets was done using credentials generated via a Google service account created on the Google Cloud Platform.
4. Unique IDs are generated for each record in Cosmos DB using a concatenation of the applicants' first name, last name, and submission date. A CosmosResourceExistsError exception class was used to accommodate for multiple records submitted by one applicant on the same date. If a duplicate ID exists during record creation, an exception is thrown by Cosmos DB. A new unique ID is then generated using exception handling.
5. Records containing a subset of fields needed for processing are added to the surrender administrator's Google sheet by an Azure function.
6. Records are mapped to the Cosmos DB document-oriented data model and loaded to Cosmos DB.

Phase 2 Validation, Extraction, and Updating to Cosmos DB

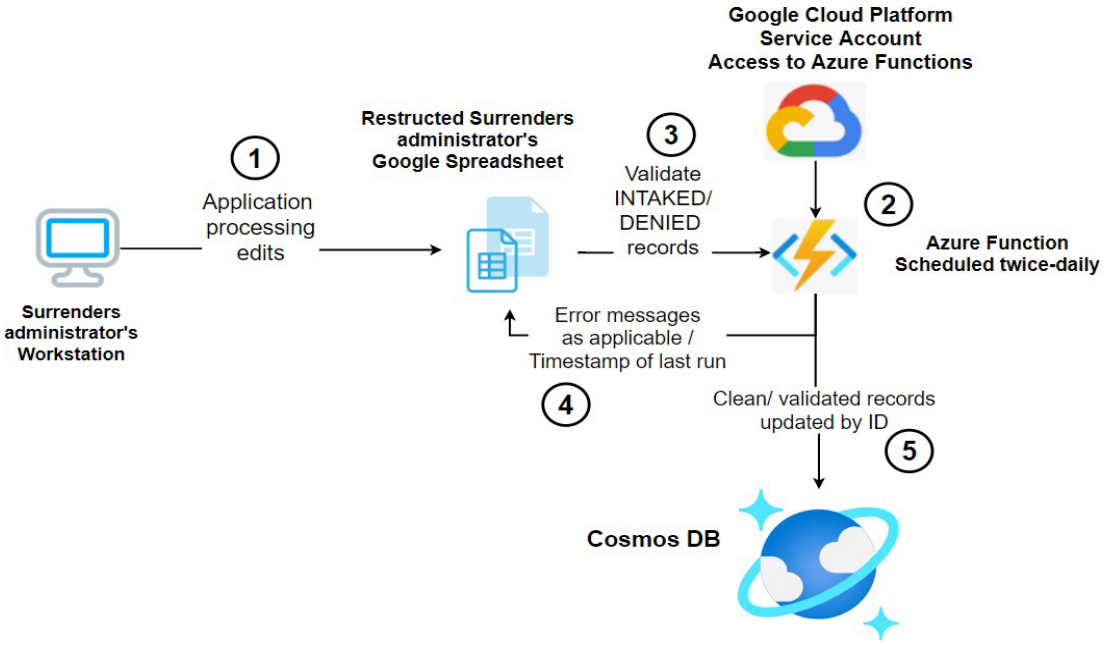

Figure 3. Phase 2 validation, extraction and updating to Cosmos DB

1. Updates are added to a surrender application by the surrender administrator.
2. Data extraction in phase two occurs on a twice daily scheduled basis.
3. Data validation is a key step of phase two before records are updated in Cosmos DB. Data validation is carried out by an Azure function. The records to be extracted are validated using validation rules identified from Secondhand Hounds' surrender application policies and procedures. Apps Script functions were also used to provide other layers of data integrity automation.
4. Error messages are generated during the validation step. These messages are added to a designated tab on the surrender administrator's sheet.

5. Records containing errors are excluded from extraction. Validated records are updated to Cosmos DB by matching the IDs of extracted records to the destination record to be updated.

Centralized storage for the surrender application on Azure cloud platform

While considering DBs, we went through the pros and cons of an RDBMS, and we had the following points in mind: 1) During requirements gathering with the Secondhand Hounds team, we found out that schema of the data stored can change from time to time as they add or remove columns depending on the need. This is a major limitation for RDBMS as due to a fixed schema and different relations between tables. It is difficult to add or remove columns from an RDBMS database. 2) An RDBMS might give scaling issues in the future with a large increase in data which Secondhand Hounds was already facing in recent years compared to past experience. If the amount of data increases by a drastic amount in the case of an RDBMS, it would require scaling up resources which might be very expensive especially for a non-profit organization with cost being a big factor (Pandora FMS team, 2022). 3) Information can be lost if incoming data is unable to be stored due to its size and storage limitations. 4) Performance might be affected during peak hours. 5) A NoSQL DB would satisfy the current use cases along with the volume of data, within the free tier offerings of AWS, and Microsoft Azure, the two cloud service providers considered.

Microsoft Azure's Azure Cosmos DB was selected as the centralized storage for the implementation. The following points were factored in, to make this selection: 1) Microsoft Azure provides $3500 worth of credits in the free tier service towards Non-profits. These credits can be applied to all Azure services (Msftnonprofits) 2) It is scalable instantly and has global availability. 3) It gives quick access and response to data queries because of fast reading and writing of data. (Nia, 2021) 4) It can be easily accessed through an API. Data requests can be facilitated using a few lines of code. 5) It has low latency and therefore is best suited for automation and fetching or updating data from the front end. 6) It has easy to understand documentation. 7) Azure cosmos DB has a flexible schema, and it is easily extensible to accommodate changing business requirements. For example, after our full-scale implementation, the surrender administrator requested to drop a column from the database. It took us limited time and effort to achieve that objective. This would have been very difficult if we had an RDBMS, as a column can be present in multiple tables and might have multiple relations.

Cosmos DB's Provisioned Throughput option was chosen over its Serverless capacity mode to take advantage of its free tier offerings. The Provisioned Throughput's Standard Scaling option was chosen to allow for more operational control.

Search capability to find any surrender application from the database

The Search capability UI uses 6 fields as identified by the Secondhand Hounds Team and as shown in figure 4.

Figure 4. Search capability UI with search fields, "Fetch records" button and the "Clear results" Button

The "Search by" option in the search UI includes searching records using separate user inputs in the fields for "Last Name", "email", "Pet ID", "Pet Name" and "Year+Month". These options function as filters for

searching records. To protect the fields from their A1 notation getting changed due to future changes in the UI, we used the concept of named ranges for all the fields required for our code. The named ranges are variables assigned for each field or range of fields' A1 notation and these variables represent a field irrespective of where they are. After the fields were assigned the named ranges, we wrote the code to fetch the user entered values from the named ranges and passed them in the specified SQL queries. Search inputs were converted to lowercase as Cosmos DB is case sensitive. We then mapped the "Search" tab columns to the keys of the items in Cosmos DB so that the correct data gets populated under the correct column. An Azure function is triggered when the "Search" button is clicked via an App Script function added to the button. The function uses the URLFetchapp service to make an Azure Functions API call. The Azure Function then passes an SQL query to Cosmos DB and also populates the sheet with results, if matching records are found.

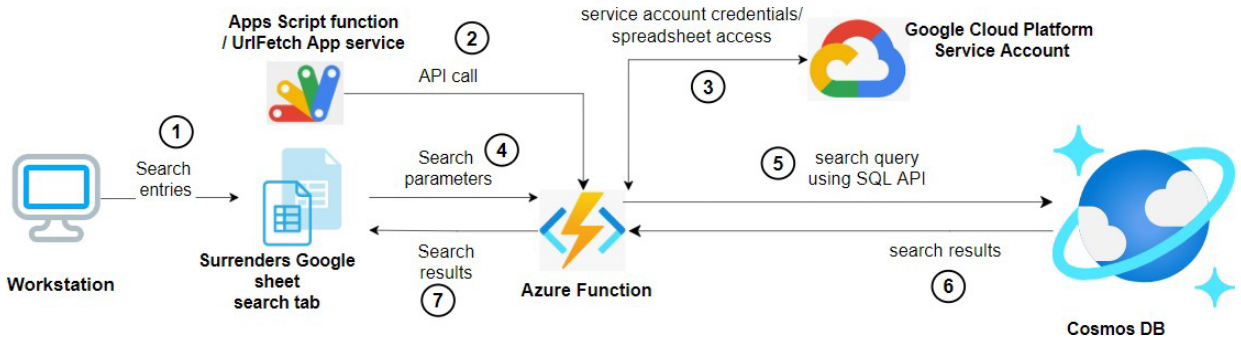

Figure 5. Search implementation using Cosmos DB SQL API

Other features of the search UI include a "Clear" button for clearing populated search results, a timestamp of the latest search carried out, and a count of the records returned from the database in a search result.

Results and findings

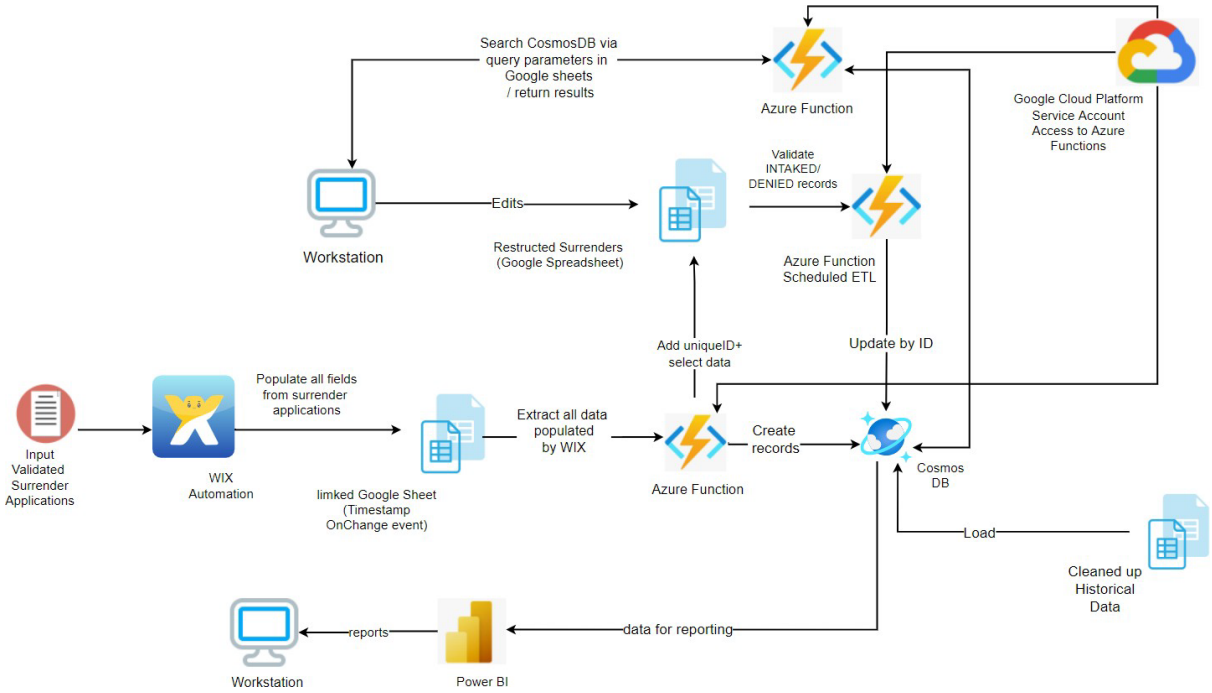

Figure 6. Fully implemented data extraction, transformation, loading, updating, and reporting solution diagram

List of technologies used:

1) WIX Automation 2) Python programming language 3) Microsoft Serverless Azure Functions 4) Google sheets API 5) Google Service Account on Google cloud platform 6) Power BI for BI reports and dashboards 7) Microsoft Azure Cosmos DB

Analytics Reports:

The analytics reports examined the surrender application data of Secondhand Hounds over the period of 2014 to 2022 (current year). Note: The data from 2017-September 2020 was missing from the historical data. To make data-driven decisions to mitigate surrenders, it was important to identify the different contexts in which the data from surrender applications can be presented by reports.

Findings and Discussions:

1. ***Number of Surrender Applications by Year:*** This report provides insight into the Number of Surrender Application on a yearly basis

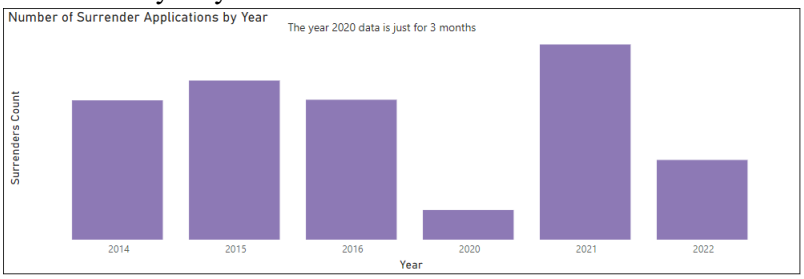

Figure 7. Number of Surrender Applications by Year

2. ***Surrender Application by Status by year:*** The year-on-year trend for number of surrender applications which were denied over the past 6 years is as given in figure 8.

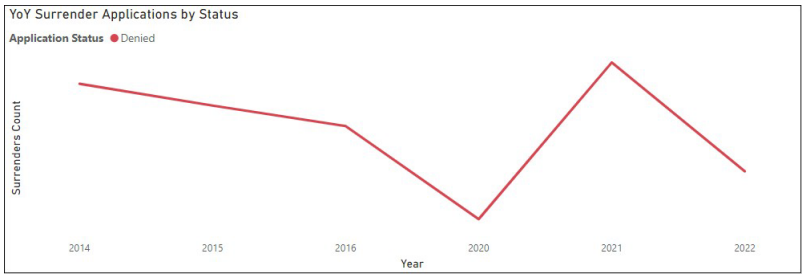

Figure 8. Surrender Application by Status by year

3. ***Number of Surrenders in various categories:*** Surrenders over the past 6 years in various categories like by Species (Dog, Cat. Pig, Rabbit), by Age Range of Pet (age of pets are categorized in 3 – Less than 1 year, between 1 to 6 years and more than 6 years), by Size of pet (Large & Bully, Cat, Fighter Fund) and by Application Status (Denied) is as shown in figure 9.
4.

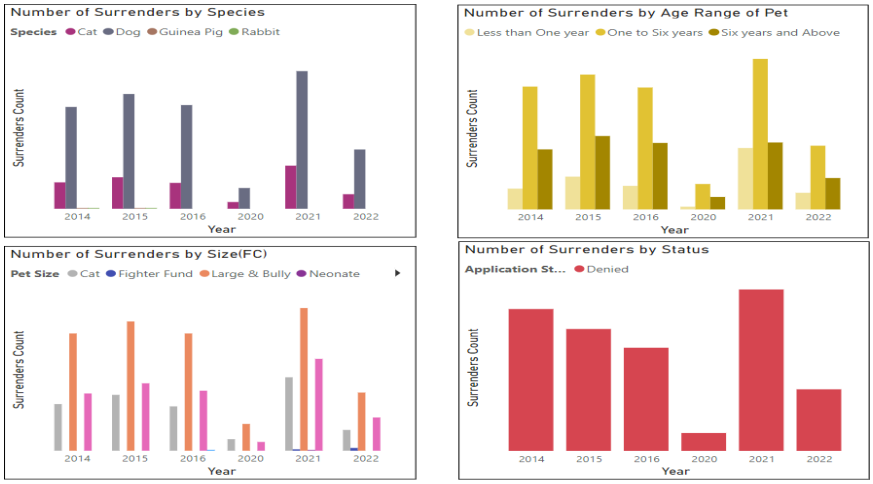

Figure 9. Number of Surrender Application by Status in various categories

5. ***Number of Adoption Returns by Year:*** Adoption returns are those pets/animals which were adopted from Secondhand Hounds and surrendered back to them. This report helps to understand the trend of these adoption returns as shown in figure 10. When the reasons for surrender are factored into an analysis of the adoption return numbers, it may be possible to get an insight into what drives adoption returns in the first place.

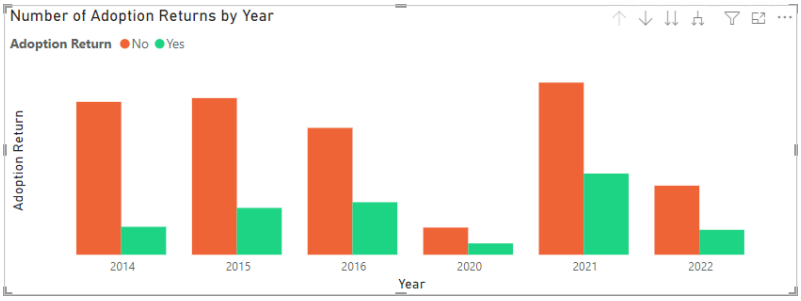

Figure 10. Number of Adoption Returns by Year

6. ***Top Reasons for Surrenders and Denial:*** The top 5 reasons for Surrenders and Denials are as shown in figure 11. This is another important report since this gives an idea on what factors Secondhand Hounds should focus on to mitigate the rise in surrenders. Also, the top reasons for denial chart provides a good insight for strategies to have fewer denials, to prevent mission drift of the animal rescue's goal to find permanent adoption homes for animals at need.

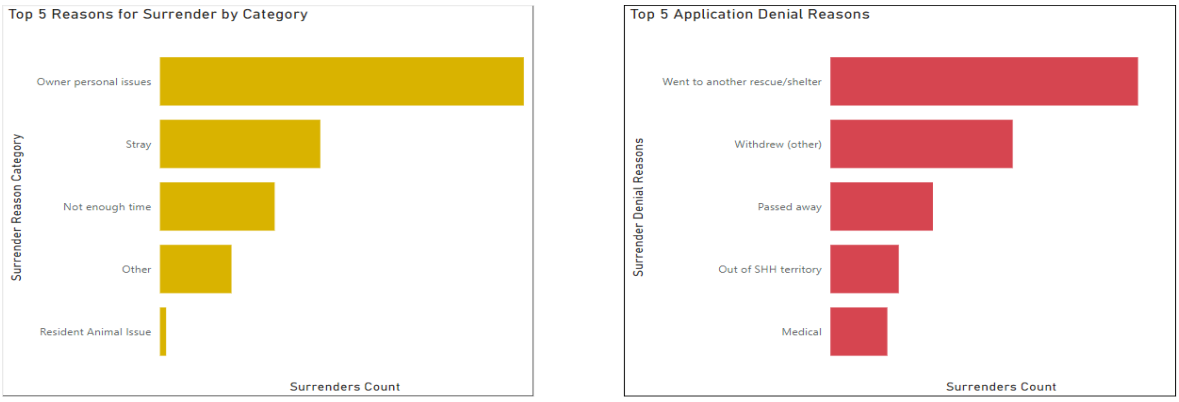

Figure 11. Top Reasons for Surrenders and Denial

Challenges and Limitations

A limitation of this solution approach is that it has been tailored to the data source for surrender applications. This data source relies on the automation feature of Secondhand Hounds' WIX website for populating a linked Google sheet with data. Changing the WIX platform would require a different approach to extract surrender applications. Another limitation is the use of Google sheets as an intermediate storage for applications that are undergoing processing. This approach could lead to unintended changes to data due to the sheet's volatility. Tools and technologies used were selected solely to make the cost of operation essentially free within the limits of Microsoft Azure's free tier service and a Microsoft Azure non-profit grant. Retrieving records from Cosmos DB using the search capabilities comes at a cost which Microsoft Azure cloud platform calculates as request units. This led to partly limiting search fields inputs in the design of the search user-interface.

Conclusion

Key areas of the current work process have been automated using Python based serverless Azure Functions and Google Apps Script functions. Populating data on the surrender administrator's spreadsheet which used to be manually carried out is now an automated process. Error prone records which had no validations are now being validated for data integrity. Errors in data entries can now be identified for corrections to be made as necessary to ensure the veracity of reports. Historical data previously stored and manually searched on Excel spreadsheets and Google sheets, are now stored and queried via a central storage on a secure platform. This central storage provides data replication and backup and an easily extensible structure for changing business requirements using a NoSQL document-oriented database. Secondhand Hounds is now able to see trends in their historical data using Power BI. These all together, in line with the aim of this project, have created an active automated collection and analysis of data for insights which have the potential to be leveraged for data-driven strategies to mitigate pet surrender applications.

Future enhancements

Future enhancements have been identified that could address some of the limitations with this solution

approach. A web application using Microsoft Azure's App Service to replace the surrender administrator's Google sheets will address the volatility of using Google sheets as an intermediate storage while surrender applications are being processed. Prediction-based reporting to inform strategic planning to mitigate pet surrenders, use of a webhook to trigger extraction, and an accommodation for changes to the data source are other areas for future enhancements.

References

Dinata, Y. M., Engel, M. M., & Antonio, T. (2021, September). *Cloud computing based data integration for NON PROFIT ORGANIZATION: Web design and implementation*. DSpace Home. Retrieved May 15, 2022, from https://dspace.uc.ac.id/handle/123456789/3994

Eid, M., Al-Jabri, I., & Sohail, M. (2018). Selection of Cloud Delivery and Deployment Models: An Expert System Approach. *International journal of decision support system technology,10 (4),17-32.* https://doi.org/10.4018/IJDSST.2018100102

Gailey, G., Batra, S., Berry, D., Buck, A., Berry, D., & Valavala, P. (2022). *Quickstart: Create a function in Azure with Python using Visual Studio Code.* Microsoft. Retrieved March, 3, 2022, from https://docs.microsoft.com/en-us/azure/azure-functions/create-first-function-vs-code-python

Glass, S., & Palmer, T. (2021, December 24). *Animal shelters see surge in pet surrenders*. NewsNation. Retrieved May 14, 2022, from https://www.newsnationnow.com/morninginamerica/animal-shelter-pet-surrenders/

Google. (2021, June 16). *Event Objects.* Google Apps Script. Retrieved March 15, 2022, from https://developers.google.com/apps-script/guides/triggers/events

Ho, J., Hussain, S., & Sparagano, O. (2021, May 7). Did the COVID-19 pandemic spark a public interest in pet adoption? Frontiers. Retrieved May 20, 2022, from https://www.frontiersin.org/articles/10.3389/fvets.2021.647308/full

Kline, A. and Dolamore, S. (2020). Understanding Data-Driven Organizational Culture: A Case Study of Family League of Baltimore. *Journal of Technology in Human Services, 38(3), 247-270.* https://doi.org/10.1080/15228835.2018.1564412

Koparkar, P., & MacKrell, D. (2015). How Fluffy is the Cloud?: Cloud Intelligence for a Not-For-Profit. *Australian Conference on Information Systems 2015, Adelaide.* https://doi.org/10.48550/arXiv.1606.00752

Longman, Green, Longman, and Roberts. (n.d.). *The AWS Nonprofit Credit Program*. Amazon. Retrieved May 14, 2022, from https://aws.amazon.com/government-education/nonprofits/nonprofit-credit-program/

Maxwell, L.N., Rotz, D., & Garcia, C. (2016). Data and Decision Making: Same Organization, Different Perceptions; Different Organizations, Different Perceptions. *The American Journal of Evaluation, 37(4), 463-485.* https://doi.org/10.1177/1098214015623634

Microsoft (2022a). *Azure.* Serverless Computing: An Introduction to Serverless Technologies. Retrieved March 15, 2022, from https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-serverless-computing/

Microsoft (2022b). *Azure.* Azure Functions Documentation. Retrieved March 22, 2022, from https://docs.microsoft.com/en-us/azure/azure-functions/

Microsoft (2022c). *Azure.* Azure Functions Pricing. Retrieved March 21, 2022, from https://docs.microsoft.com/en-us/azure/azure-functions/pricing

Msftnonprofits. (n.d.). *Azure cloud solutions: Microsoft for Nonprofits*. Microsoft. Retrieved May 14, 2022, from https://www.microsoft.com/en-us/nonprofits/azure?SilentAuth=1&wa=wsignin1.0

Nia, N. K. (2021, May 21). *Pros and cons of Azure Cosmos DB 2022 - Trustradius*. Retrieved May 15, 2022, from https://www.trustradius.com/products/azure-cosmos-db/reviews?qs=pros-and-cons

Pandora FMS team. (2022, February 18). *NOSQL vs SQL: Key differences and how to choose*. Pandora FMS Monitoring Blog. Retrieved May 14, 2022, from https://pandorafms.com/blog/nosql-vs-sql-key-differences/

Weiss, T., Chen, D., Eissing, R., Assaf, W., Gunda, S., Raghupathy, R. K., & Watford, C. (2022, April 12). *Azure cosmos DB free tier*. Microsoft Docs. Retrieved May 14, 2022, from https://docs.microsoft.com/en-us/azure/cosmos-db/free-tier