TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Ella Werthan 184071IVSB

A Corporate Monitoring Solution for CCPA and GDPR Compliance: the Case of Lyft

Bachelor's thesis

Supervisor: Valdo Praust

Master of Science,
Cybernetics

Tallinn 2021

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Ella Werthan 184071IVSB

ETTEVÕTTE JÄLGIMISE LAHENDUS CCPA JA GDPR-I JÄRGIMISEKS: LYFTI JUHTUM

bakalaureusetöö

Juhendaja: Valdo Praust

Magistrikraad,
küberneetika

Tallinn 2021

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Ella Werthan

[14.05.2021]

Abstract

In recent years, global trends towards broad data privacy rights, such as the right to be forgotten, are accelerating. In the coming years, vast numbers of global citizens will gain these rights, and companies will need to make plans to bring their operations into compliance with the associated regulations. This thesis is a case study of the work done at a Californian company, Lyft, to comply with the California Consumer Privacy Act and provide its users with a right to the erasure of their personal data.

The work done in this thesis represents a successful implementation of a compliance monitoring strategy in the Support Operations organization at the company. This monitoring system has contributed to Lyft's successful efforts to achieve best practices and avoid financial and reputational penalties associated with noncompliance.

This thesis is written in English and is 23 pages long, including 7 chapters and 4 figures.

Annotatsioon

Ettevõtte jälgimise lahendus CCPA ja GDPR-i järgimiseks: Lyfti juhtum

Lõputöö on kirjutatud Inglise keeles ning sisaldab teksti 23 leheküljel, 7 peatükki, 4 joonist.

List of abbreviations and terms

API	Application Programming Interface
CCPA	California Consumer Privacy Act
ETL	Extract, Transform, Load
GDPR	General Data Protection Regulation
IPO	Initial Public Offering
NASDAQ	National Association of Securities Dealers Automated Quotations
PII	Personally Identifiable Information
RR	Resolution Reason
SR	Submission Reason

Table of Contents

1 Introduction....9
2 Background....11
2.1 CCPA....11
2.2 GDPR and Right to Erasure....13
2.3 Lyft....13
2.4 Support Operations....14
2.5 Technologies....15
3 Problem....16
4 Process....18
4.1 Gathering Information....18
4.1.1 Policies and Procedures....19
4.1.2 Ticketing Software....20
4.1.3 Areas of Concern....20
4.2 Creating the Data Source....21
4.3 Driving Visuals....25
5 Results....27
5.1 Efficacy of the Monitoring Dashboard....27
5.2 Insights from the Data Source....28
6 Conclusion....30
7 Summary....31
References....32
Appendix 1 – Non-exclusive licence for reproduction and publication of a graduation thesis....34

List of Figures

Figure 1: Relevant Tables and Fields.....22
Figure 2: Key Ticket Snapshots.....23
Figure 3: Linking Account Deactivations to Support Interactions.....24
Figure 4: Resolution Reasons for Data Privacy Tickets Resulting in Deactivations.....29

1 Introduction

As the provision of services to individual consumers has been increasingly digitized over recent decades, legal discussions surrounding data collection and retention have become a hot-button issue among digital policy experts. Vast amounts of personal data are now integral to the operations of a majority of companies, and this data can be collected in increasingly covert ways. While these covert methods of data collection have facilitated more streamlined user experiences, they have allowed consumers to release personally identifiable information (PII) without their knowledge. Data privacy experts are now advocating for requiring for entities to obtain user consent prior to collecting PII and allowing for users to withdraw this consent at their will.

This so-called "Right to be Forgotten" has been a well-known component of European digital policy for several years following the entry into force of the General Data Protection Regulation (GDPR) in 2016. While other areas of the world have been slow to adopt such broad data privacy rights, legislation like the California Consumer Privacy Act (CCPA) in the United States has increased the number of consumers with the right to be forgotten by tens of millions.

Re-working policies, procedures, and technologies to account for ongoing user consent takes time and effort. Solutions likely will not function flawlessly following initial implementation. Given these considerations, companies handling consumer data would be wise to proactively implement the right to be forgotten, as well as a comprehensive monitoring strategy to remediate gaps in compliance. This thesis is an explanation of the background, process, and results of a project intended to accomplish this at a large, publicly traded technology company in the United States.

As a part of her work in the summer of 2020, the author of this thesis was tasked with the creation of a monitoring system for CCPA and GDPR compliance among the host company's third party contracted customer support associates. This monitoring system would focus on the handling of data deletion requests and the common confusion

among support associates between data deletion and user deactivation. This monitoring system would be driven by a new data source that the author would create, connecting previously unrelated data tables using complex business logic.

This thesis will first explain the background elements of the work conducted, including relevant information regarding the host company, data privacy regulations, and business intelligence data concepts. This information will provide context for the type of work conducted and will bolster the reader's understanding of the problem at hand.

The author will then give a thorough explanation of the problem solved by the work in this thesis. This section will give a clear idea of what would define a successful solution to the problem. The author will then detail the work done to solve this problem, from information gathering to authoring code.

Finally, the author will explore the results of this work. This section will summarize the work done, analyse the efficacy of the solution the work provided to the problem at hand, and extrapolate any additional value of the work.

2 Background

This project was carried out at Lyft, Inc., a software company in the United States. It was undertaken in preparation for the enforcement of CCPA, a piece of data privacy legislation specific to the state of California. The focus of the thesis will be a monitoring solution for their Support Operations team for compliance with the data privacy elements of CCPA, as well as adherence to internal policies and procedures. These background elements will be explored in the following sections for greater understanding of the problem and solution described in this thesis.

2.1 CCPA

The California Consumer Privacy Act (CCPA) is a sweeping data privacy law enacted in 2018 governing the data privacy rights of the State of California's 39.5 million residents[1]. It applies to any entity earning over $25 million in gross annual revenue, processing the personal data of over 50,000 California residents, or deriving more than half of their annual revenue from the sale of personal data belonging to California residents[2].

From the website for the Office of the Attorney General of the State of California[3], key rights defined in the CCPA include the following:

> The right to know about the personal information a business collects about them and how it is used and shared;
>
> The right to delete personal information collected from them (with some exceptions);
>
> The right to opt-out of the sale of their personal information; and

1 See [1]
2 See [2]
3 See [2]

The right to non-discrimination for exercising their CCPA rights.

The project described in this thesis will focus on the right of California residents to delete personal information collected from them.

The exact language of these "data deletion" provisions[1] relevant to this thesis is as follows:

> (a) A consumer shall have the right to request that a business delete any personal information about the consumer which the business has collected from the consumer.
>
> (c) A business that receives a verifiable consumer request from a consumer to delete the consumer's personal information pursuant to subdivision (a) of this section shall delete the consumer's personal information from its records and direct any service providers to delete the consumer's personal information from their records.

While the CCPA was enacted in 2018, the law came into effect on 1 January 2020 and enforcement was delayed until 1 July 2020[2]. Enforcement of the CCPA is carried out by the Attorney General of the State of California. Consumers can report failure to comply by filing a complaint with the Attorney General's office. If the Attorney General finds that the entity in question is indeed non-compliant, the entity is given a confidential notice of the finding and a 30-day window for remediation[3].

If sufficient remediation actions have not been completed in this 30-day window, the Attorney General will recover fines by filing a public lawsuit against the entity in question. The fines for non-compliance are $2,500 for each unintentional violation of the law and $7,500 for each intentional violation of the law[4].

1 See [3]
2 See [4]
3 See [2]
4 See [4]

2.2 GDPR and Right to Erasure

This project's host company, Lyft, as described below, provides a physical service and does not operate within the European Union. Thus, it is not currently subject to the wide-ranging data privacy provisions of the GDPR. However, Lyft is a fast-growing company with eventual expansion ambitions. While dedicating time and resources to achieve compliance with CCPA, it made sense to consider those portions of the GDPR that mapped to the CCPA in preparation for potential future operations within the European Union. The compliance department at the host company wrote the following description of the combined requirements of the GDPR and the CCPA as they relate to data privacy and the Support Operations organization:

> CCPA and GDPR both require that we maintain a process to receive, analyse, and respond to user requests to exercise their privacy rights. This includes requests to delete or access their information.

Thus, when GDPR compliance is referenced in this thesis, it is restricted to this requirement.

2.3 Lyft

Lyft, Inc. is a technology startup founded in 2012. It is publicly traded on the National Association of Securities Dealers Automated Quotations (NASDAQ) exchange under the symbol LYFT and was valued at $24.3 billion in its initial public offering (IPO) in 2019[1]. The company provides a ride-sharing platform connecting drivers and passengers with a mobile application and operates within the United States and Canada. Lyft has corporate offices in dozens of cities and partners with several third party customer support providers.

There are two types of application users in the Lyft platform: drivers and riders. Both types of users must submit verified personal information for safety reasons. Additionally, drivers are subject to mandatory background checks. As of 2019, the Lyft mobile application had been downloaded over 65 million times[2]. This means that Lyft

1 See [5]
2 See [6]

processes tens of millions of users' personal and potentially sensitive data for normal business functions.

With California being the largest state in the United States, and having been founded there, a significant share of Lyft's user base are protected by the CCPA. Additionally, it is expected that many other states, and eventually the nation, will enact similar regulations. It was thus a business necessity to reach compliance.

The CCPA's method of enforcement provided enhanced incentive for a functioning CCPA compliance plan. A lawsuit brought by the Attorney General against a publicly traded company would be widely reported by mainstream media outlets. California's residents were set to vote on a resolution in November 2020 to decide whether Lyft's fundamental business model of non-employed drivers could continue to operate. A CCPA lawsuit would have dealt a devastating blow to the company's public image and could have been enough to sway the results of the ballot measure.

2.4 Support Operations

Support Operations at Lyft is the internal organization responsible for developing, implementing, governing, and monitoring processes for providing application end users with support. The organization consists of many teams, including the Support Analytics team, which the author of this thesis belonged to, and Support Quality, a group of stakeholders that the Support Analytics team produces work for.

The Support Analytics team consists of a group of Business Intelligence analysts who produce work for other teams within the Support Operations organization. These analysts use data collected at various endpoints within internal applications and third party platforms to provide insights and data-driven solutions for issues relating to user support.

The Support Quality team is a group of specialists tasked with ensuring quality support is provided to users by the support associates directly providing assistance. These support associates are a combination of internal team members and employees of third party vendors. The support associate teams are spread across dozens of global cities.

2.5 Technologies

The Support Operations organization at Lyft uses a variety of technologies to process, track, and analyse data arising from customer support interactions. The lifecycle of a customer interaction begins with a ticketing software called Zendesk. Over the course of the interaction, important metrics are collected from the Zendesk endpoint. After an interaction has closed, the data points collected from that interaction undergo an Extract, Transform, and Load (ETL) process using Apache Hive in order to be stored as usable audit logs.

While Hive is ideal for the ETL process due to its storage capabilities, the Support Analytics team additionally uses the Presto Query Engine to perform more targetted analysis on the support data. This is due to the fact that Presto is generally considered to exhibit better performance for ad-hoc interactive queries[1].

These Presto queries used in targetted support data analysis are developed and stored in Mode, a data science platform. While not explicitly intended for complex query development, Mode is a great platform for these efforts because it is collaborative, version controlled, and easily integrated with R and Python. Additionally, queries in Mode can be extended with Liquid for analyses requiring more elaborate logic[2].

Also at the Support Analytics team's disposal is an internal, open source tool called Amundsen. Named from the first person to discover the South Pole, Roald Amundsen, Amundsen is an integrated metadata library. Amundsen allows analysts to explore vast numbers of data tables and provides points of contact for the owners of these tables[3]. With immeasurable data points collected at Lyft each day, Amundsen allows analysts and stakeholders to navigate to the best tables for their use cases.

1 See [7]
2 See [8]
3 See [9]

3 Problem

Prior to the author's work at Lyft, the internal compliance team had developed policies and procedures for compliance with CCPA requirements for responding to data deletion requests. These policies and procedures included robust documentation for user support associates to refer to when support tickets are flagged as relating to data privacy.

Several security experts, including Deloitte, an internationally-known consulting firm, place a heavy emphasis on training when it comes to ensuring employee compliance with data privacy controls.[1] The author of this thesis found this advice to be lacking. A robust training and awareness campaign is important, but Deloitte's recommendations neglect the importance of monitoring noncompliance. Other data privacy voices, such as IBM[2] and industry expert JC Cannon[3], acknowledge the advantage in monitoring employee behaviours for compliance.

The author of this thesis decided that compliance monitoring was a necessary component of Lyft's data privacy compliance programme. With hundreds of types of support tickets and dozens of support offices around the world, Lyft could apply more targetted remediation with a data-driven approach to compliance monitoring. The author chose to leverage the data collected in existing audit logs of the support ticketing and user management systems to monitor incidences of CCPA and GDPR noncompliance.

The next priority in assuring CCPA compliance was a monitoring system for ensuring that these processes were being followed and for identifying problematic scenarios relating to the handling of data privacy requests. The author was tasked with creating this monitoring system using existing user data from endpoints in Lyft's support ticketing platform, Zendesk, as well as Lyft's user account portal.

1 See [10]
2 See [11]
3 See [12]

The ideal monitoring system would show aggregate trends in data privacy requests and break down these trends by third party support vendor and ticket lifecycle. It would highlight trends of noncompliance and allow the end user to drill down to individual tickets for further investigation. The monitoring system would provide a supplement and a guide for annual manual audits of data privacy support tickets.

4 Process

To create a data privacy compliance monitoring system, the author would first need to fully understand the procedures that support associates were directed to follow. This would allow for a better understanding of the behaviours that would indicate noncompliance. The author would then review the results of prior manual audits of data privacy tickets. These results would provide an initial sense of where issues with compliance were arising. Finally, the author would observe support associates using Lyft's ticketing platform, Zendesk, to understand the meaning behind the data points collected from the platform.

After the author gathered enough information to plan the monitoring system, the building efforts would begin with the creation of a single data source to drive a monitoring dashboard. This data source would join points from numerous data tables and endpoints to capture information for each ticket that bore relevance to data privacy compliance.

Finally, the author would use the information gathered regarding support procedures and the data source created with that information to drive an interactive visual dashboard of adherence to data privacy policies within the Support Operations organization. This dashboard would need to update regularly with the latest data and would be designed with Support Quality Specialist end users in mind. This dashboard would be used to regularly monitor the organization's CCPA compliance and would serve as a supplement to annual manual audits of data privacy tickets.

4.1 Gathering Information

The initial stage of this project was to gather information from the Support Operations organization that would aid the author of this thesis in constructing the data privacy compliance dashboard. The author would need to know the policies created by the internal compliance team to bring the organization into compliance with CCPA. The

author would also need to understand how support associates used Lyft's ticketing software in order to accurately use data points collected from the software to track compliance. Finally, the author would need to review the results of manual audits to understand the scenarios that the Support Quality team was interested in tracking.

4.1.1 Policies and Procedures

To gather information on the internal policies at Lyft governing data privacy issues and CCPA compliance, the author was directed to read documents in an internal library referred to as Compass. The policies that Lyft's support associates follow in responding to any user ticket are found here. Compass contains interactive documentation that allows support associates to input user answers to their questions to view the correct course of action for each possible scenario.

The support tickets relating to data privacy, CCPA, GDPR, and account deletion issues link to a Compass article titled "Data Privacy." While assisting users, support associates would use the guiding questions provided in the article and input any user answers along the way. The Compass article would guide the associate along one of three paths:

1. Redirection. If the support associate determines that a user is requesting to delete their account and data or download the data that Lyft has collected for their account, the support associate will be directed to redirect the user to the data privacy portal in the Lyft mobile application to make such requests formally. The support associate would make this redirection by providing links, directions, and explanations of the entire process.

2. Education. If the support associate determines that a user is requesting further information about Lyft's data privacy policies, the support associate will be directed to provide the user with education about these policies. The materials in Compass for user education include Lyft's policies regarding data privacy, such as what data Lyft collects, who has access to it, and whether Lyft is compliant with CCPA and GDPR.

3. Escalation. If the support associate determines that a user is referencing a threat to legal action, has a complex issue, or is requesting more specific information than the materials in Compass can provide, the support associate will be directed to escalate the ticket to be handled by an authorized agent.

If the support associate determines that none of these three paths is appropriate and the ticket has been mistakenly flagged for data privacy, the associate will apply the correct flag to the ticket and navigate to an appropriate Compass article.

4.1.2 Ticketing Software

Support associates at Lyft use a help desk ticketing software called Zendesk for handling user requests. Since the author would be using support data from Zendesk to drive the data privacy compliance monitoring dashboard, it was decided that the author would shadow a support associate to learn more about what each data field represented in the ticket lifecycle.

Each ticket comes to the assigned support associate with a Submission Reason (SR). This submission reason is a flag for type of ticket that links to a corresponding Compass article for instructions on resolution. The support associate communicates with the user by asking the questions recommended by the Compass article.

When a solution is reached within the Compass documentation, the support associate will apply this solution in Zendesk by selecting the corresponding Resolution Reason (RR), which provides a predetermined response for the user. The support associate will then close the ticket.

Each correspondence the support associate makes with the user is recorded in a log called a "verbatim." The verbatim, in combination with SR, RR, user identification, support associate identification, and timestamps, provides auditable data of the assistance provided and is collected to the Lyft internal databases through a Zendesk application programming interface (API) endpoint. This information is what the author of this thesis would use to drive the data privacy compliance monitoring dashboard.

4.1.3 Areas of Concern

To gain context and wisdom on the scenarios involving data privacy that had previously arose concern, the author consulted with the Support Quality team. This team is in charge of ensuring that the support associate teams and third party providers are providing assistance in accordance with Lyft policies and standards. They had been

tasked with manual audits of tickets for data privacy policy compliance in the past, and would be the target users of the monitoring dashboard.

In these consultations, the author learned that their main area of concern was account deactivations. Account deactivations differ from account deletions in that Lyft retains the user data associated with deactivated accounts. Further, user data deletion is handled through a portal accessible through the end user mobile application. So, if a user requested an account deletion through the help desk, and a support associate mistakenly deactivated the user's account, the user's data would not be deleted and the user would have no way to request data deletion. This would be clearly noncompliant with both CCPA and GDPR.

To complicate this matter, support associates are not asked to record account deactivations in Zendesk, and are also not asked for a Zendesk ticket number in the user portal when deactivating the account. This means that there would be no key in either the deactivation record or the ticket record to link the two events.

The Support Quality team had found numerous cases of mistaken data privacy related account deactivations in their manual audits. However, without a way to link the two events and pull aggregate data, they had no way to make data-driven decisions about where to apply remediation actions and what that remediation should look like. They would need information on the trends of which third party teams were making this mistake and what kinds of tickets were most likely to result in a mistaken deactivation.

4.2 Creating the Data Source[1]

Before writing any queries, it was important to illustrate clearly which pieces of information would be useful to capture in the monitoring dashboard. The author of this thesis began by using the Amundsen metadata system to identify which tables and fields would be needed to successfully track an interaction from the support ticketing system to the user management system. A summary of the relevant fields and tables identified by the author is displayed below:

1 Although this section references developing a query, it will not be possible to replicate this query here due to an agreement the author signed legally designating the query itself as the intellectual property of Lyft. The author has made an earnest attempt to explain and illustrate the logic of the query without showing any code.

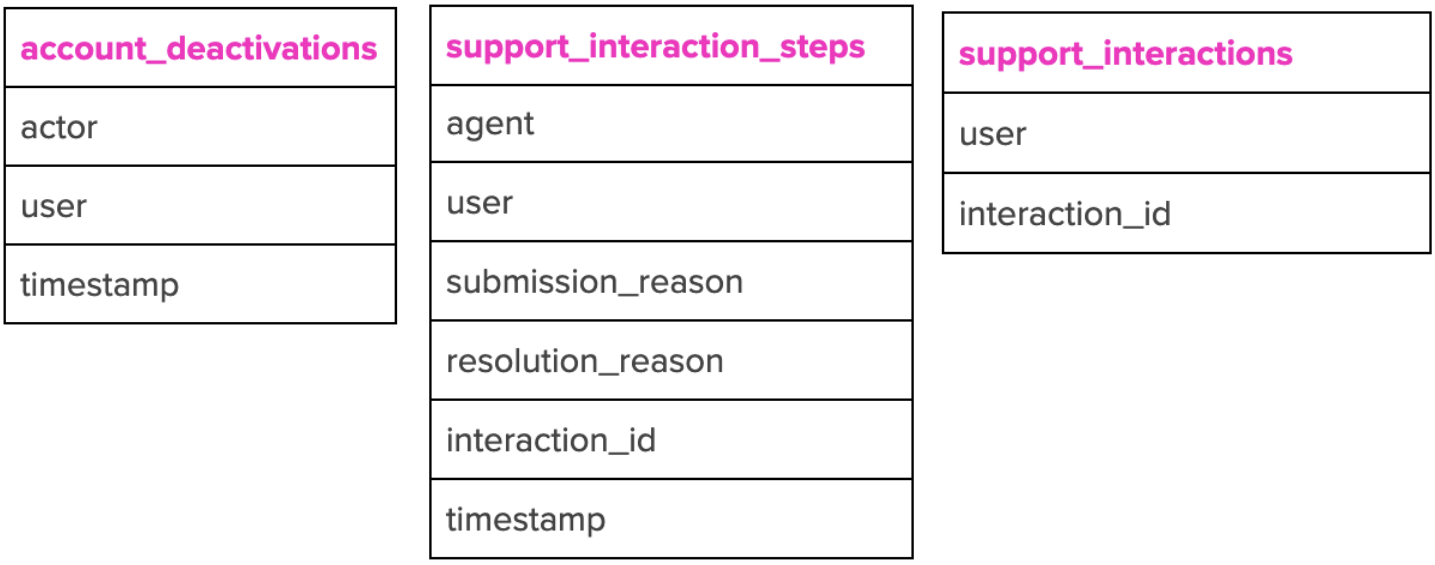

Figure 1: Relevant Tables and Fields

In the internal databases, each support ticket is recorded as a record in a Support Interactions table. Additionally, each modification of a ticket over the course of its lifecycle generates a snapshot record in a Support Interactions Steps table. Since tickets can be modified any number of times in their path to resolution, it made sense to create a standardized set of snapshots to collect from each support ticket represented in the dashboard.

The figure below represents these standardized steps. The Submission Reason of the Support Interaction Steps record immediately preceding a user deactivation would represent which Compass procedure the support associate would have been directed to follow. The Resolution Reason of the Support Interaction Steps record immediately following a user deactivation would represent how the support associate described their own actions. The initial and final states of the ticket are captured in each Support Interaction record.

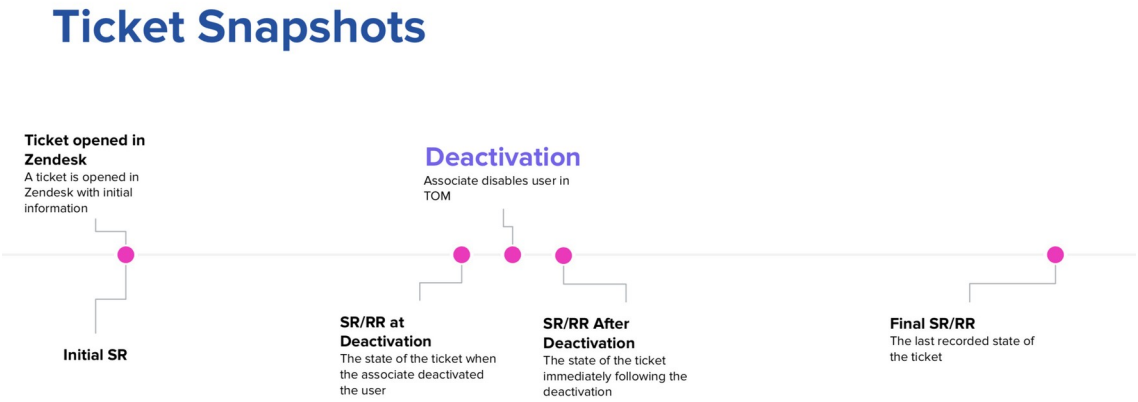

Figure 2: Key Ticket Snapshots

With an understanding of these key snapshots, the author could begin developing logic for linking support tickets to user deactivations. As shown in Figure 1, the support interactions and support interaction steps tables are linked with a key, interaction_id. However, the interactions tables and user_deactivations table are not. The author of this thesis had to develop query logic to link these records using detailed knowledge of the tables and the support ticket lifecycle.

The Support Quality and Support Analytics teams were most interested in deactivations that occurred when support associates should have been following the Data Privacy procedures in Compass. These would be Support Interactions with the most recent step preceding an associated user deactivation having a Submission Reason of "Data Privacy."

Therefore, the first step to linking data privacy tickets to user deactivations used the more granular Support Interaction Steps table. The author started by writing Presto code extended with Liquid in the Mode analytics platform to filter this table to records with an SR of "Data Privacy". Then, the author performed an inner join on these records and the user deactivations table using the following conditions:

1. The actor who deactivated the user account was the same as the support associate assigned to the ticket.

2. The deactivated user was the user who initiated the support ticket.

3. The deactivation happened after the record in the Support Interaction Steps table.

This code underwent a peer-reviewed Quality Assurance process to verify the results. In this process, the author of this thesis found that there were some cases in which multiple steps in an interaction joined to one user deactivation as a result of the joins in the code. The goal was to end up with a one-to-one relationship between deactivations and interactions. To pull only the *latest* step in an interaction preceding the user deactivation, the author wrote additional code to select the largest timestamp for records with each represented Support Interaction ID. These results represented the "Step Before" and "Deactivation" in the timeline in Figure 3.

The initial and final states of the relevant tickets were then included by joining the records to the Support Interactions table on Support Interaction ID. To complete the records, the author also selected the earliest Support Interaction Steps records following the deactivation associated with the Support Interaction ID in the step record.

A visual summary of these final results is represented in Figure 3 below.

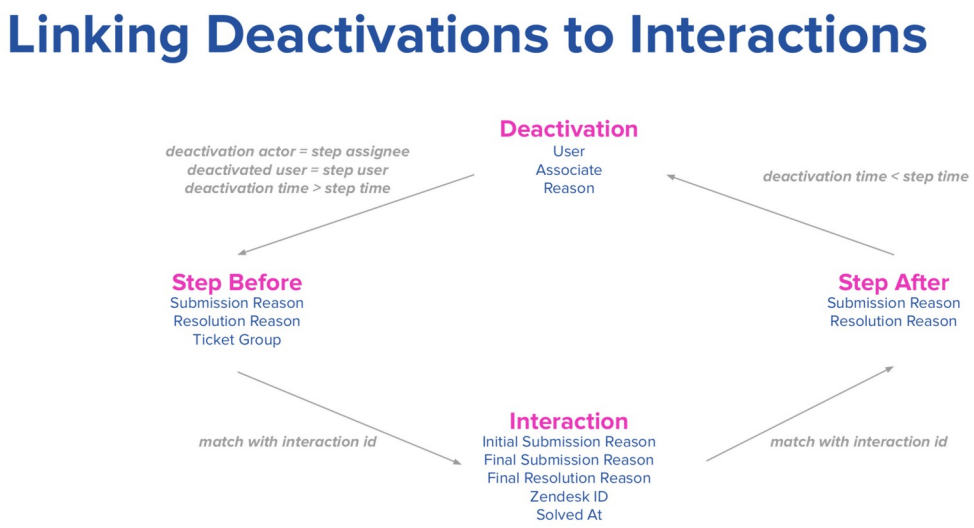

Figure 3: Linking Account Deactivations to Support Interactions

4.3 Driving Visuals[1]

With the data source created, it was time to use it to drive visuals in an interactive monitoring dashboard. To do this, the author used a Business Intelligence tool called Tableau. The Tableau visual dashboard was connected to the internal database using the query the author designed as a custom parameter. The author consulted with the Support Quality team to determine how often they planned to use the dashboard and created a data refresh schedule that was short enough to prevent stale data but long enough to avoid unnecessary query costs.

In deciding which visuals to include, the author returned to earlier discussions with the Support Quality team. A key part of their job was coordinating with Lyft's third party support partners regarding areas for improvement. Given this fact, the author included a graph that showed data privacy deactivations by support partner and location. This way, the Support Quality team could determine which partner organizations and specific locations were experiencing difficulties maintaining CCPA compliance and remediate accordingly.

It was also important to the Support Quality team to have a way to monitor their remediation efforts. If, for example, they spent time and resources developing new Data Privacy documentation in Compass, it would be important to be able to see if there were resultant reductions in the number of data privacy related deactivations. To that end, the author created a chart of the number of such deactivations over time. The partner and location chart described above acted as a filter for this chart – if a user clicked on a specific partner or location, they could see the number of deactivations over time for that specific group of support associates. This would allow for monitoring of more targeted remediation efforts such as in-person trainings.

The author included an interactive visual showing the lifecycle of data privacy support tickets associated with deactivations. This would assist the Support Quality team in identifying which types of tickets are eventually classified as data privacy tickets and where confusion may arise. This visual allowed users to see percentages of tickets in the data source with each type of initial Submission Reason, final Submission Reason, and

1 Similar to the query the author coded, the dashboard itself is the intellectual property of Lyft and contains user data that cannot be shown in this paper. Instead, the author focuses on the reasoning behind her decisions to design each visual included in the dashboard

Resolution Reason. By clicking on any given Submission Reasons or Resolution Reasons, the subsequent columns would only account for records filtered by the selected category. This visual also included the percentage of the total data source represented by the filtered selection.

Finally, the author included an exportable spreadsheet of all of the tickets selected using the filters on the sheet. This would allow the Support Quality team to quickly and easily generate populations for auditing and would facilitate detailed drill-downs into specific scenarios.

5 Results

As a result of the work conducted in this thesis, the author provided the host company with two resulting assets:

1. A Business Intelligence dashboard designed to monitor noncompliance with CCPA-driven procedures governing data deletion requests.

2. A data source connecting support ticket events and account deactivation events. This data source allows the Support Quality team to monitor and audit user deactivations without relying on self-reporting from support associates.

The author of this thesis measured the efficacy of the data privacy compliance monitoring dashboard both by gathering data during the initial presentation of the tool in the summer of 2020 and by monitoring CCPA enforcement trends in the following months. In conversations within the Support Operations organization at Lyft, the Support Quality and Support Analytics teams also determined a few key insights regarding the support-driven deactivations data source that would help drive further improvements in the support procedures. By analysing support tickets linked to deactivations using the query designed by the author, it because clear that many of these deactivations would never have been captured using the metrics available prior to the work in this thesis.

5.1 Efficacy of the Monitoring Dashboard

The first measure that the author used to determine the efficacy of the data privacy compliance monitoring dashboard was to compare it with the results of previous manual audits for data privacy compliance. Since CCPA is enforced with fines for each individual infraction, and not organizational trends, it is important that any individual instances of noncompliance are proactively identified when possible. An indicator of the success of this dashboard would be if it could catch more serious infractions per population than the audits.

When compared with an audit from three months prior to the creation of the dashboard, which looked at a sample of the entire population of data privacy tickets, the data privacy deactivation tickets sampled by the dashboard identified twice as many cases of confirmed noncompliance with CCPA. This was a massive indicator that the dashboard would be an asset in reducing the risk of incurring CCPA fines due to failure to facilitate user data deletion.

Additionally, while the Attorney General of the State of California has filed numerous lawsuits over the course of 2020 and 2021 against companies who have not complied with CCPA, Lyft has not been sued or disciplined for noncompliance[1]. As discussed earlier in this paper, avoiding CCPA-related financial penalties was important, but perhaps more important was avoiding the public scandal that would come with noncompliance enforcement ahead of the November elections. Indeed, with the passage of Proposition 22[2], Lyft's business model was allowed to continue and its market capitalization grew by $1.8 billion[3], an eye-popping 22% growth from its capitalization before the election.

These early and persistent indicators point to the data privacy compliance dashboard created by the author of this thesis being part of a successful implementation of a CCPA compliance program.

5.2 Insights from the Data Source

In addition to monitoring for data privacy procedure adherence, the data source that the author of this thesis created for this project provided important insights for further improvements in the Support Operations organization at Lyft. Because deactivations had never been linked to support tickets of any kind previously, the author generalized the data source by removing the requirement that tickets be related to data privacy. This allowed the Support Quality team to examine trends in deactivations for any type of ticket.

1 See [13]
2 See [14]
3 See [15]

An important discovery that resulted from this exercise was that the Submission Reason that most commonly resulted in a user deactivation was linked to Compass documentation that did not direct support associates to deactivate the user. While not an issue for CCPA or data privacy compliance, this indicated a major deviation from expected behaviour. Many Compass directives, like the Data Privacy ones, are rooted in compliance with laws and regulations. It is important that associates are following these directives and that any deviations from them are tracked. This discovery highlighted the value of tying deactivations to support tickets.

Additionally, when examining a sample of the deactivations tied to support tickets, it became apparent that many support associates were not using the built-in reporting metrics to indicate a deactivation. When a support associate deactivates a user account in connection with a support ticket, the support associate is directed to add a Resolution Reason of rider_deactivation_confirmation or driver_deactivation_confirmation.

When the author of this thesis sampled a group of deactivations relating to Data Privacy tickets, there were five unique Resolution Reasons:

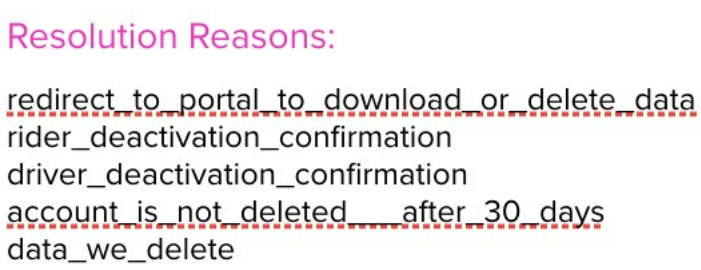

Figure 4: Resolution Reasons for Data Privacy Tickets Resulting in Deactivations

As displayed in Figure 4, the most common Resolution Reason, redirect_to_portal_to_dowload_or_delete_data, does not indicate a deactivation. In fact, three of the five Resolution Reasons in this sample group would have indicated a CCPA and GDPR-compliant handling of a Data Privacy request. However, knowing that the support associates in these cases deactivated the users in question, this indication would have been false. This outlines the importance of the data source designed by the author of this thesis and the inadequacy of self-reported metrics in audit logs used for compliance monitoring.

6 Conclusion

The CCPA directive relevant to this project was simple: maintain a process to receive, analyse, and respond to user requests to exercise their privacy rights. The path to compliance, however, was not. It involved dedicated efforts from dozens of teams across the organization. From scoping, planning, and writing policies and procedures, to implementing, monitoring, and auditing, it was an "all hands on deck" effort.

In the emerging landscape of data privacy rights, companies might be tempted to ignore regulations like GDPR and CCPA if they do not operate in the European Union or California. But all indications point to these rights, like the right to be forgotten, spreading globally[1] over the coming years. With the amount of effort required to comply with these regulations, companies that wait too long may find themselves woefully unprepared.

As is often the case in the field of cybersecurity, the monitoring project detailed in this thesis wasn't an immediate driver of revenue. But failure to ensure organizational compliance with CCPA would have come at an enormous cost. If found to be noncompliant with CCPA, Lyft would be ordered to pay potentially massive fines. Additionally, a dip in public opinion in 2020 could have forced Lyft to abandon its business model entirely, and likely would have resulted in losses in market capitalization measuring in the billions.

If consulted, the author of this thesis would recommend to any company the immediate implementation of a robust compliance plan for laws like CCPA, and that this plan include data-driven monitoring of procedures. If entities wait too long, they will be forfeiting the opportunity to improve and harden their compliance efforts prior to enforcement in their localities. This could result in dire business consequences.

1 See [16]

7 Summary

The author of this thesis was tasked by the host company with creating a monitoring solution to track noncompliance with user support procedures necessitated by the California Consumer Privacy Act (CCPA). To do this, the author first conducted thorough research of internal policies and the CCPA, including the intended outcomes of data privacy related user support tickets. Next, information was gathered regarding the typical lifecycle of a user support ticket. The author then held multiple meetings to ascertain the main areas of concern around data privacy with the Support Quality team. All of the information gathered was leveraged to design a monitoring dashboard for data privacy compliance.

To implement the designed dashboard, the author first created a query to join several tables in the database according to specifications agreed upon with the Support Quality team. These joined tables formed a new data source linking user deactivations to support tickets. Then, the data source was connected to a Business Intelligence tool to create visuals. The author worked closely with the Support Quality team to design visuals that would be useful to their mission.

The dashboard was a component of a successful implementation of a data privacy compliance program. The data source can successfully identify non-compliant behaviours that would not have been identified with self-reported data. As a result of the data source and monitoring system created by the author of this thesis, as evidenced by public records, the host organization has been able to avoid noncompliance and the associated financial and reputational consequences.

References

[1] "U.S. Census Bureau QuickFacts: California." U.S. Census Bureau. Accessed 2021. https://www.census.gov/quickfacts/fact/table/CA/PST045219.

[2] "California Consumer Privacy Act (CCPA)." State of California – Department of Justice – Office of the Attorney General, March 3, 2021. https://www.oag.ca.gov/privacy/ccpa.

[3] "California Law - Code Section." California Legislative Information, 2020. https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?sectionNum=1798.105.&nodeTreePath=8.4.45&lawCode=CIV.

[4] Gates, Megan. "CCPA Deep Dive: How California Is Enforcing Its Major Privacy Law." ASIS Online, December 1, 2020. https://www.asisonline.org/security-management-magazine/articles/2020/12/ccpa-deep-dive-how-california-is-enforcing-its-major-privacy-law/.

[5] O'Donnell, Carl, and Joshua Franklin. "Lyft Valued at $24.3 Billion in First Ride-Hailing IPO." Reuters. Thomson Reuters, March 28, 2019. https://www.reuters.com/article/us-lyft-ipo-idUSKCN1R92P4.

[6] Arevalo, Tony. "20+ Lyft Statistics — Rides, Market Share, and Revenue (2020 Edition)." Carsurance, February 24, 2021. https://carsurance.net/blog/lyft-statistics/.

[7] "Introduction to Presto (PrestoDB)." Amazon, 2015. https://aws.amazon.com/big-data/what-is-presto/.

[8] "Answer Any Question with SQL." Mode. Accessed 2021. https://mode.com/online-sql-editor/.

[9] Hall, Susan. "Lyft's Amundsen: Data-Discovery with Built-In Trust." The New Stack, July 16, 2019. https://thenewstack.io/lyfts-amundsen-data-discovery-with-built-in-trust/.

[10] "GDPR Compliance Monitoring." Deloitte, February 26, 2019. https://www2.deloitte.com/mt/en/pages/risk/articles/mt-risk-article-gdpr-monitoring.html.

[11] Gottshall, Justine, and Adam C. Nelson. "Developing a Privacy Compliance Program." IBM. Thomson Reuters, 2016. https://www.ibm.com/downloads/cas/AR34REKO.

[12] Rodriguez, Deidre. "Monitoring Your Privacy Program: Part Three." IAPP, March 24, 2015. https://iapp.org/news/a/monitoring-your-privacy-program-part-three/.

[13] "CCPA Case Tracker." O'Melveny. O'Melveny & Myers, February 16, 2021. https://www.omm.com/resources/alerts-and-publications/alerts/ccpa-case-tracker/.

[14] "California Proposition 22, App-Based Drivers as Contractors and Labor Policies Initiative (2020)." Ballotpedia. Accessed 2021. https://ballotpedia.org/California_Proposition_22,_App-Based_Drivers_as_Contractors_and_Labor_Policies_Initiative_(2020).

[15] Mohamed, Theron. "Uber and Lyft Gain $13 Billion in Combined Market Value after Californians Approve Prop 22." Business Insider, November 4, 2020.

https://markets.businessinsider.com/news/stocks/uber-lyft-stock-prices-california-votes-for-prop-22-2020-11-1029764137.

[16] Perarnaud, Clément. “Right to Be Forgotten.” GIP Digital Watch, February 12, 2021. https://dig.watch/issues/right-to-be-forgotten#view-10940-2.

Appendix 1 – Non-exclusive licence for reproduction and publication of a graduation thesis[1]

I Ella Werthan

1 Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis "A Corporate Monitoring Solution for CCPA and GDPR Compliance: the Case of Lyft", supervised by Valdo Praust

1.1 to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;

1.2 to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.

2 I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.

3 I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

[14.05.2021]

1 The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.