The Kronos Incident: Geospatial-Temporal Patterns of Life Analysis

William Tandio – Data Science Capstone Term Project

1. Project Background

In January, 2014, the leaders of GAStech are celebrating their new-found fortune as a result of the initial public offering of their very successful company. In the midst of this celebration, several employees of GAStech go missing. An organization known as the Protectors of Kronos (POK) is suspected in the disappearance, but things may not be what they seem.

Many of the Abila, Kronos-based employees of GAStech have company cars which are approved for both personal and business use. Those who do not have company cars have the ability to check out company trucks for business use, but these trucks cannot be used for personal business. Employees with company cars are happy to have these vehicles, because the company cars are generally much higher quality than the cars they would be able to afford otherwise. However, GAStech does not trust their employees. Without the employees' knowledge, GAStech has installed geospatial tracking software in the company vehicles. The vehicles are tracked periodically as long as they are moving.

This vehicle tracking data has been made available to law enforcement to support their investigation. Unfortunately, data is not available for the day the GAStech employees went missing. Data is only available for the two weeks prior to the disappearance. In addition to the vehicle data, law enforcement has been given access to the personal and business credit and debit card transactions for the local GAStech employees for the two weeks preceding the kidnapping. Many of the GAStech employees also use loyalty cards to gain discounts or extra benefits at the businesses they patronize, and law enforcement has been given access to two weeks of this loyalty card data as well.

As a data scientist expert assisting law enforcement, your mission is to make sense of this data to identify suspicious patterns of behavior and to prioritize which of these may be related to the missing staff members. You must cope with uncertainties that result from missing, conflicting, and imperfect data to make recommendations for further investigation.

2. Dataset

- **Car Assignment :**
  A list of vehicle assignments by employee, in CSV format (`car-assignments.csv`) :
  1. Employee Last Name
  2. Employee First Name
  3. Car ID (integer)
  4. Current Employment Type (Department; categorical)
  5. Current Employment Title (job title; categorical)

- **ESRI shapefiles of Abila and Kronos (in the Geospatial folder)**
- **GPS Data:**
  A CSV file of vehicle tracking data (`gps.csv`)
  1. Timestamp
  2. Car ID (integer)
  3. Latitude
  4. Longitude
- **Loyalty Card Data:**
  A CSV file containing loyalty card transaction data (`loyalty_data.csv`)
  1. Timestamp
  2. Location (name of the business)
  3. Price (real)
  4. FirstName (first name of the card holder)
  5. LastName (last name of the card holder)
- **Credit Card Data:**
  A CSV file containing credit and debit card transaction data (`cc_data.csv`)
  1. Timestamp
  2. Location (name of the business)
  3. Price (real)
  4. FirstName (first name of the card holder)
  5. LastName (last name of the card holder)
- A tourist map of Abila with locations of interest identified, in JPEG format (`map-tourist.jpg`)

3. Project Goals

1. Describe common daily routines for GAStech employees. What does a day in the life of a typical GAStech employee look like?

2. Identify up to twelve unusual events or patterns that you see in the data. If you identify more than twelve patterns during your analysis, focus your answer on the patterns you consider to be most important for further investigation to help find the missing staff members. For each pattern or event you identify, describe
   - What is the pattern or event you observe?
   - Who is involved?
   - What locations are involved?
   - When does the pattern or event take place?
   - Why is this pattern or event significant?
   - What is your level of confidence about this pattern or event? Why?

3. Like most datasets, the data you were provided is imperfect, with possible issues such as missing data, conflicting data, data of varying resolutions, outliers, or other kinds of confusing data. Considering data is primarily spatiotemporal, describe how you identified and addressed the uncertainties and conflicts inherent in this data to reach your conclusions in questions 1 and 2.

4. Data Cleaning and Preprocessing

The data preparation process for the project involved several crucial steps to ensure the integration and cleaning of disparate data sources. This section outlines the key aspects of data preparation, including the merging of datasets, data cleaning, and feature engineering.

1. **GPS Data**
   The primary dataset used in the analysis is the GPS data, which was combined with the car assignment data to create a comprehensive dataframe. By merging the "gps.csv" file with the "car_assignment.csv" file, a cohesive dataset was created. This integrated dataframe provides essential information such as timestamps, longitude, latitude, driver names, car IDs, and employment details. This dataset serves as the foundation for mapping employee locations at specific points in time, enabling the visualization of their movements two weeks prior to the incident.
2. **Stop Data**
   Given that the GPS data lacks explicit stop information, the "stop data" fills this gap. A specialized function was developed to iterate through the main dataframe, calculating time intervals between rows. When the time interval exceeded five minutes, a stop flag was incorporated into the dataset, enabling a comprehensive understanding of where employees visited during their days.
3. **Card Data**
   Card transaction data provide insights into employees' financial activities. The integration of credit card and loyalty transaction data into a unified dataframe facilitated a comprehensive analysis of financial behavior. Transaction type flags was added to distinguish between credit card and loyalty card transactions, enabling a nuanced exploration of spending patterns. This combined dataset is created to unravel potential correlations between employee behavior and financial activities, further strengthening the analytical framework.

These python files collectively laid the base groundwork for the dashboard development using Plotly Dash. The processed merging, cleaning, and categorization of diverse data sources created in a unified dataset ready for analysis. This integrated dataset helps this project to examine and bring together employee geospatial-temporal patterns with credit card transaction records. The dashboard, a visual interface powered by the prepared data, used as an important tool to analyze employee movements, investigate anomalies, and establish accuracy in credit card transactions.

5. Data Visualization

The visualization phase of the project used Plotly Dash to provide an interactive and comprehensive representation of the cleaned and merged data. The integration of Plotly Dash component tabs facilitated a structured and user-friendly interface for exploring the diverse facets of employee movement and behavior.

Truck Driver Movement Tab:

The first tab, dedicated to "Truck Driver Movement," addresses the unique challenge of identifying truck drivers for various routes and dates. The dynamic graph displayed on this tab enables users to select specific routes and dates. By correlating the GPS data with the truck routes, this visualization aids in pinpointing the responsible truck driver for each route on distinct occasions. This insight provides clarity into truck driver assignments and route allocation, contributing to a comprehensive understanding of the geospatial-temporal patterns associated with truck drivers.

General Employee Movement Tab:

The second tab introduced an intuitive dropdown menu enabling the selection of distinct employee categories. An additional time dropdown facilitated the isolation of movements during specific periods of the day. The chosen parameters dynamically updated the graph, allowing users to visualize employee movement patterns based on their category and chosen time. Complementing the graph, a table provided details, enhancing the understanding of employee trajectories. A weekend radio button was added, to switch between weekday and weekend views. This option aimed to unveil disparities in general employee movement behaviors between these timeframes.

Individual Employee Movement Tab:

In the "Individual Employee Movement" tab, users can select an employee's name, along with a date and time range. The resulting graph provides a detailed trajectory of the selected employee's movement during the specified period. A supplementary table showing card transaction data is positioned under the graph. This integration facilitates a comprehensive analysis, enabling users to discern potential anomalies or irregularities in an individual employee's movement patterns and associated financial activities. By comparing movement data with card transactions, this tab supports the identification of suspicious behaviors and aids in confirming the legitimacy of transactions.

6. Project Analysis

6.1 The life of a typical GAStech employee

The life of a typical GAStech employee, as revealed through the dynamic visualizations of the Plotly Dash dashboard, offers a fascinating glimpse into their daily routines and spatial preferences. This comprehensive analysis underscores the interplay between employment type, geography, and daily habits, weaving a narrative that paints a vivid picture of their lives.

Figure 1 : Employees resident segmentation

Firstly, the dashboard highlights a remarkable trend where employees of similar employment types tend to cluster in specific residential areas. Engineers, for instance, predominantly reside along Carnero Street, while executives have made Spetson Park their home, and the IT team is concentrated in the western region, nestled between Arkadiou and Sannan Park. This spatial alignment points to a sense of community and camaraderie among colleagues with shared professional backgrounds.

6.1.1 Morning

Figure 2 & 3 : Western vs Eastern Resident go to Coffee Shop

In the morning those residing in the eastern area embark on their commute around 7:00 AM, while employee that lives in the western area leisurely set out at 7:45 AM. Executives, known for their early-bird tendency, commence their day at an even earlier 6:30 AM. The first place that employees generally visit in the morning is a coffee shop, they get their morning coffee from places closest to their homes.

6.1.2 Afternoon / Lunch

Figure 4: Afternoon / lunch time

At lunchtime, around 12 PM, GAStech employees disperse from the office to various destinations. The area around Epsilon Avenue witnesses a bustling gathering, becoming a favored spot for many employees. Security personnel patrolling the town's perimeters. Ouzeri Elian visited by some, while others opt for convenience, visiting grocery stores, marts, Robet & Sons, or Albert's Fine Clothing. This midday breaks not only show diverse preferences but also provides a snapshot of the employees' choice of place during their break.

6.1.3 Evening

Figure 5 & 6 : Western vs Eastern Resident after hours pattern

As the working hours ends, GAStech employees follow distinct departure schedules, with executives leaving later around 6:15 PM and others at approximately 5:30 PM. Evening routines reveal a blend of activities, particularly for those residing in the west. Many head home briefly before going out for dinner near Ipsilantou Avenue. Notably, individuals living in the west, especially executives, occasionally visit Albert's Fine Clothing before retreating for the day. Lucas Alcazar stands out as a regular customer of Ouzeri Elian. These evening insights highlight the employees' varied after-hours habits and engagements.

6.1.4 Night

The night to early morning timeframe remains relatively dormant, characterized by minimal activity. Any deviations from this norm are flagged as anomalies, emphasizing the stability of this period in the lives of GAStech employees.

6.1.5 Weekend

Figure 7 : Executive Sunday golf

Executives engage in golfing from afternoon till evening, enjoying a recreational escape. During daytime every employee has a span of diverse activities without any distinctive pattern. However, as night falls, a trend can be seen - those residing in the west frequently visited Hippokampos, while their eastern counterparts lean towards Ipsilantou Avenue. This geographical distinction displays night activities preferences, providing a summary into how different residential locations weekend evening socializing choices within the GAStech community.

6.2 Truck Driver Pattern

Analyzing the general patterns of truck drivers within GAStech presents a unique challenge, given the absence of specific driver-truck assignments. However, the utilization of the dashboard in addition with transaction data has shown a method to solve the responsible driver, associated truck, and their weekly routes. Through comparison and cross-referencing of the dashboard's graph with individual transaction records, the table of truck driver schedule can be created.

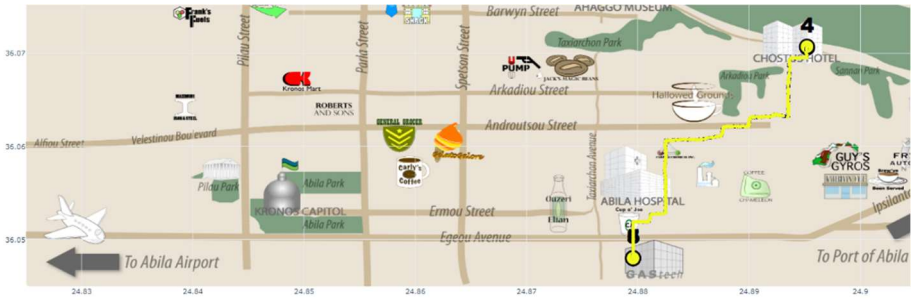
Table 1 : Truck Driver Routine

Truck ID	Operating Day	Day	Driver	Description
101	Mon,Tu,Thu,Fri	Mo	Albina Hafon	
		Tu + Thu	Benito Hawelon	
		Fri	Claudio Hawelon	Not Operating on 17th
104	Mon-Thu		Henk Mies	Only Airport
105	Tue-Thu		Valerie Morlun	
106	Tue-Thu	Tue+Thu	Dylan Scozzese	
		Wed	Adan Morlun	
107	Mon-Fri		Cecilia Morluniau	Morning - Lunch
		7,8,9,15 Jan	Irene Nant	After Lunchtime

7. Anomaly

7.1 Insignificant to Investigation Anomalies

- Who : Isande Borrasca and Brand Tempestad
  When : Lunch time on January 8,10,14,and 17
  Where: Chostus Hotel



Card Transaction Table

timestamp_full	FullName	location	price
2014-01-08 13:19:00	Brand Tempestad	Chostus Hotel	111.89
2014-01-08 12:00:00	Isande Borrasca	Brewed Awakenings	12.32
2014-01-08 12:56:00	Isande Borrasca	Chostus Hotel	107.51

Figure 8 : Isande and Brand Hotel Lunch

These 2 person are the only employees that went to Chostus Hotel during lunch time, possible explanations is that they are having an affair or eating at the hotel restaurant. Their credit card transactions shows that each person is being charged, thus the later hypothesis might hold true.

- Who : Lars Azada, Adra Nubaron, Axel Calzas, Birgitta Frente, Brand Tempestad, Gustav Cazar, Isak Baza, Isande Borrasca, Felix Balas, Lidelse Dedos, Linnea Bergen, Lucas Alcazar, Marin Onda, Nils Calixto, Vire Frente
  When : Friday Night, January 10
  Where : Lars Azada Home

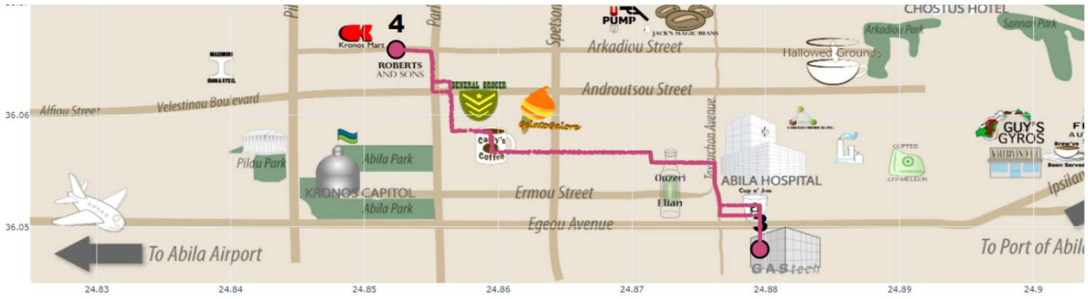
  Since its Friday night, its probably a house party held by Lars. Most employees that participated are Gastech engineers and IT.

- Who : Lucas Alcazar
  When : Night, January 6,8 and 15
  Where : Gastech Building

Figure 9 : Lucas overtime route

Lucas often works late, sometimes he had to go back to the office late at night even though he already left in the evening.

- Who : Linnea Bergen and her team
  When : Friday Afternoon, January 17
  Where : Kalami Kafenion



Card Transaction Table

timestamp_full	FullName	location	price
2014-01-17 13:16:00	Linnea Bergen	Kalami Kafenion	140.78
2014-01-17 12:00:00	Isak Baza	Jack's Magical Beans	7.57

Figure 10 : Linnea Bergen lunchbreak with the team

Linnea Bergen had a spike on her credit card transaction with $140.78 being charged to her card but her GPS location was not updated anywhere during the time of the transaction. After investigating the whole IT team, no one had any transaction at that time beside Linnea. Linnea might be treating lunch for the team as she is the IT group manager. This theory is supported with GPS tracker showing that Nils Calixto stopped at Kalami Kafenion. The team might be carpooling in Nils car's that day. Since the median transaction at Kalami Kafenion is at $26.245, lunch for 5 person (number of people in IT) around 140 dollar make sense (including tips and tax).

- Who : Kanon Herrero, Loreto Bodrigi , Edvard Vann, Elsa Orilla, and Adra Nubarron
  When : Saturday Afternoon, January 18
  Where : Kronos Capitol

Figure 11 : Some employee went to Kronos Capitol on weekend

There could be an event happening at Kronos Capitol during this time, some of people from the engineering security teams are here.

- Who : Henk Mies
  When : Night, January 16
  Where : Abila Airport

  Henk Mies usually gets back at the Gastech building before 5pm, but only on this date that he starts leaving the airport around 5pm and returns the truck past 9pm. This is the only truck that returned to Gastech building at night. Further investigation to Hank Mies card transaction didn't show anything suspicious.

- Who : Several Employee
  When : Weekdays
  Where : Mostly Coffee Shop or Restaurant

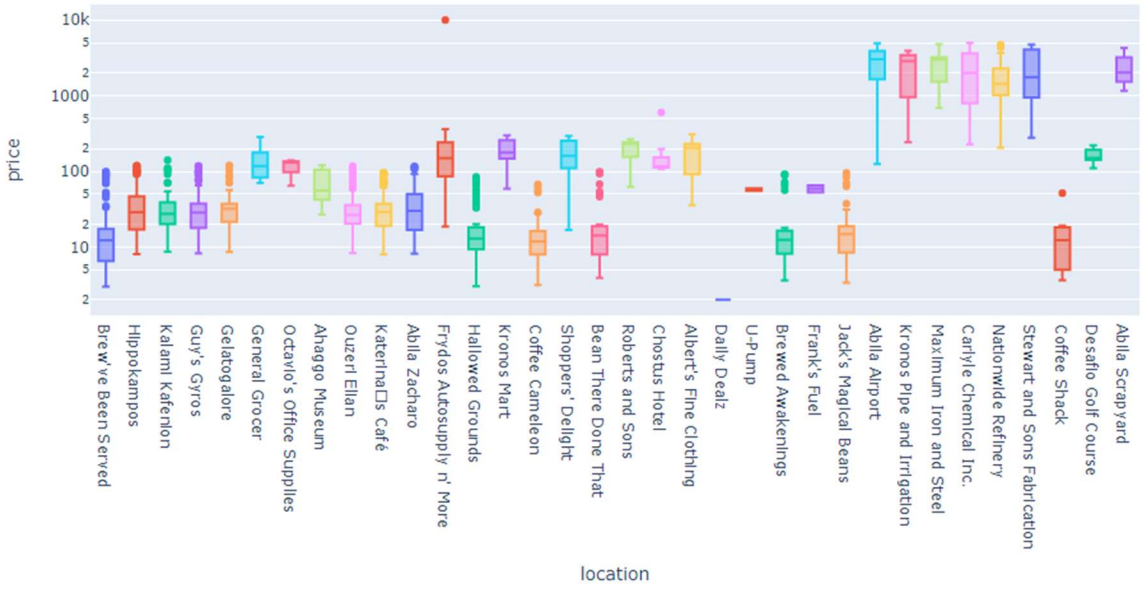


Figure 12 : Outliers in several location

There are several outliers in shops transaction, especially coffee shop and restaurant. While it is normal to have a wide range of bills charged into a credit card in a restaurant, some outliers are very far from the median value. Some outliers could be identified and explained with further investigation with graph visualization, but only assumptions can be made regarding the transaction. The most common reason for these outliers would be employees are buying / treating food / coffee for other employees.

7.2 Significant to Investigation Anomalies

- Who : Loreto Bodrogi , Isia Vann, Minke Mies, Hennie Osvaldo
  When : Midnight – Early Morning | 7,9,11 and 14 January
  Where : Executive Houses (Ada Campo Corrente, Orhan Strum, Willem Vasco Pais, and Ingrid Barranco)

Anomalies from the graphed general movement of the employee have shown a group of individuals who needs further investigation: Loreto Bodrogi, Isia Vann, Minke Mies, and Hennie Osvaldo – all of whom hold positions within GAStech's security division. Intriguingly, their activities have been notably conspicuous during the midnight to early morning hours on specific dates: January 7th, 9th, 11th, and 14th. What sets these instances apart is the fact that their GPS locations seem to converge at the residences of prominent executives, namely Ada Campo Corrente, Orhan Strum, Willem Vasco Pais, and Ingrid Barranco. Such patterns raise suspicions, suggesting potential ulterior motives. Given the context of a kidnapping investigation in this project, it is possible that these security personnel could be linked to the abduction. Their seemingly suspicious actions and location near to executive houses could indicate a deeper involvement, possibly implicating them as suspects or, at the very least, individuals of interest in a kidnapping scenario.

Table 2 : Suspect Spying Shift

Who	When	Where
Loreto Bodrogi & Isia Vann	7-Jan	Ada Campo Corrente House
Minke Mies and Loreto Bodrogi	9-Jan	Orhan Strum House
Isia Vann and Hennie osvaldo	11-Jan	Willem Vasco Pais House
Minke Mines and Hennie Osvaldo	14-Jan	Ingrid Barranco House

- Who : Isia Vann & Loreto Bodrogi
  When : Morning, 7 Jan& 14 Jan
  Where : Near Ada Campo Corrente House Area

After spying Ada Campo Corrente House the whole night, the morning Ada campo left the house to the office on January 7th, she was being followed by Loreto & Isia to the place she always visited everyday before going to work. The place that she visited is unknown, as there is no transaction happening every time she stops at this place and the map didn't show a specific detail about this location, but it is located east of Coffee Shack. Beside Ada Campo, Ingrid Barranco and Isande Borrasca also visits this place every morning. This event happened again on 14th Jan where Minke Mies went to this location after spying on Ingrid house during that night, but unlike Loreto and Isia he arrived a bit later.

Figure 13 :Loreto & Isia Following Ada Campo

Ada Campo Corrente followed by Minke after he's done spying on Ingrid

- Who : Loreto Bodrogi, Minke Mies, Inga Ferro, and Hennie Osvaldo
  When : Generally Afternoon on Weekday
  Where : Several Mysterious Place

Figure 14 : Suspicious Locations

The name mentioned above often visit these 5 mysterious places where there is no detail of what location it is. They generally visited these places during the lunch break and there are no transactions on their card being charged during their visit to these places. Again the names are similar to the security member that spies on the executive houses but Isia Vann is not included here, instead there's Inga Ferro who is also the member of security team. The table below describes their visiting dates and area.

Table 3 : Records of Members from security team that visit these suspicious places

Name / Area	1	2	3	4	5
Loreto Bodrogi	7,8,11,14	8,17	15	9	13
Inga Ferro	9	17,18	13,15	7,10	16
Minke Mies	10,16	8	9,14	15	7
Hennie Osvaldo	16	9	15	10,17	8,11

- Who : Inga Ferro, Loreto Bodrogi, Hennie Osvaldo and Isia Vann
  When : Everyday
  Where : Loreto Bodrogi / Isia Vann / Inga Ferro House

Figure 15 : Possible POK Basecamp

These people seems to live together, as the always start their day and go home to the same GPS location everyday except for Hennie Osvaldo who went to a stay at a different place on weekends. Since these people are very suspicious, this location could be an accommodation for POK.

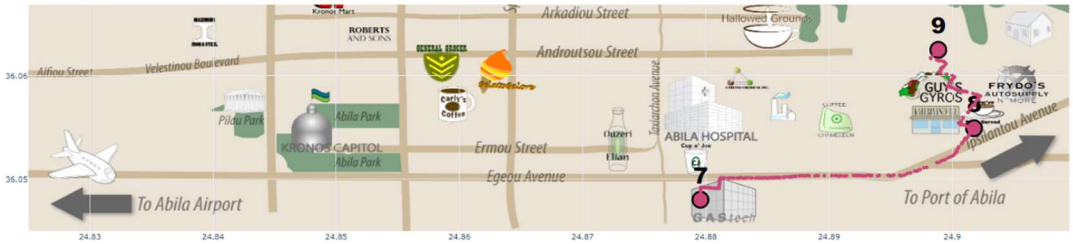
- Who : Lucas Alcazar & Minke Mies
  When : Night, 13 Jan
  Where : Frydos Autosupply n' More

On January 13th night Lucas's credit card was charged $1000 at 19:20, although his GPS location showed that he was not present at the auto supply establishment during that time; his whereabouts were instead traced to Ouzeri Elian. Several potential explanations surface, each with its own complexities. Furthermore, after the suspicious activities done at night by the security team, further investigation focused on those security members. One theory points to

potential credit card theft by Loreto Bodrogi, given his proximity to Ouzeri Elian that night and move to near his home which is coincidentally near from Frydos Autosupply n' More.

Figure 16 :Loreto Location on the night of 13th

An alternative hypothesis implicates Minke Mines, whose GPS location more accurately aligns with Frydos Autosupply n' More. Yet, no corresponding transaction from Minke Mies corroborates this theory. This hypothesis also creates suspicion towards Henk Mies the truck driver, as they shared family name and with a big purchase of a $1000, it might be items that need to be transported by truck. However further investigation showed that Henk never drives past / near the shops, thus he is out of suspicion.



Card Transaction Table

Figure 17 : *Minke Mies Location on the night of 13th*

Possible credit card theft happened supported with the fact that Lucas's credit card sees no further usage following this event, limited only to his loyalty point utilization up to the 16th Jan.

8. Addressing the Uncertainties and Conflicts

The process of addressing uncertainties and conflicts within the spatiotemporal dataset involved a strategic combination of visualization techniques and cross-referencing to ensure conclusions. Employing Plotly Dash for comprehensive data visualization facilitated a clearer understanding of the conflict present. By comparing the GPS tracking history of each vehicle with employee transactions, a clear picture can be obtained. For example, the challenge of unassigned trucks was tackled by graphing truck routes for specific dates with corresponding truck driver card transactions, effectively identifying relations.

To handle outliers, a boxplot visualization method was employed. This technique not only highlighted potential anomalies but also provided a means for deeper investigation. Suspicious GPS tracking instances, indicating possible irregularities, were subjected to thorough investigation and further exploration. For example, Lucas's suspicious credit card transaction was found thought the outlier in the boxplot, and further investigated using data visualization and data analysis.

When addressing missing or conflicting data, leaving it as it is approach was adopted. Instances of missing data were not uniformly deemed as bad data, as they often held hidden insights. Some columns with missing data were subjected to graphing analysis, revealing patterns that would otherwise remain hidden. The decision to retain certain missing data was informed by the recognition that their absence could hinder the understanding of the broader scope of events in Abila.

In essence, the synergy of visualization, comparison, and exploration of missing or conflicting data harmonized to reveal complexities within the dataset. This comprehensive approach create the accuracy of conclusions drawn and ensured a strong foundation for making informed conclusion about the spatiotemporal patterns and anomalies present in the GAStech dataset.