Project Proposal
VAST Challenge 2021 MC-2

Team Members

Manan Gandhi - mgandh10@asu.edu
Nihar Thakkar - njthakkar@asu.edu
Devam Sheth - dmsheth1@asu.edu
Niketan Chandarana - nchandar@asu.edu
Tirth Hihoriya - thihoriy@asu.edu
Hitansh Shah - hshah79@asu.edu

Problem Statement

The purpose of this analysis is to investigate the disappearance of several employees during a natural gas production company's IPO celebration. The company's track record for environmental responsibility is not favorable, and there is suspicion that a local environmental organization may be involved in the disappearance. The analysis will examine data related to employee car tracking, credit card transactions, and loyalty card usage that occurred before the incident.

This project is centered around Mini-Challenge 2 (MC2) of the VAST Challenge 2021, which is based on a made-up story about the disappearance of employees of a company. Our goal is to perform a variety of tasks using the provided datasets, including identifying common locations where credit and loyalty cards were used and detecting any unusual patterns in their usage. We will also examine anomalies in the data after adding information about vehicle movements. Additionally, we will attempt to determine the owners of each credit and loyalty card, considering evidence and uncertainties. Another task is to identify any possible informal relationships that exist among the company's personnel. Finally, we will try to pinpoint any suspicious activity and determine its location.

Proposed Implementation

Domain Abstraction

Problem: Investigation into the disappearance of several employees during a natural gas production company's IPO celebration

Intended audience: Investigators, law enforcement officials, and other parties involved in the investigation. The visualizations are designed to help answer the key questions of interest and provide insights into the potential causes of the disappearance of the employees.

Datasets to be used:

1. Vehicle assignments (car-assignments.csv)
2. GPS tracking data (gps.csv)
3. Loyalty card transaction data (loyalty_data.csv)
4. Credit and debit card transaction data (cc_data.csv)
5. Tourist map (MC2-Tourist.jpg)

Data Abstraction

This section shows the dataset given and various attributes of each dataset table.
Data:

1) Car-assignment.csv : Tabular

Attribute	Type	Scale
Last name, First name	Categorical	45

CarID	Categorical	45
CurrentEmploymentType	Categorical	5
CurrentEmploymentTitle	Categorical	21

2) Cc_data.csv: Tabular

Attribute	Type	Scale
Timestamp	Ordered(Quantitative, Sequential)	13-14 days
Location	Categorical	34
Price	Ordered(Quantitative, Sequential)	10000
Last 4 creditcard number	Categorical	55

3) Gps.csv: Tabular

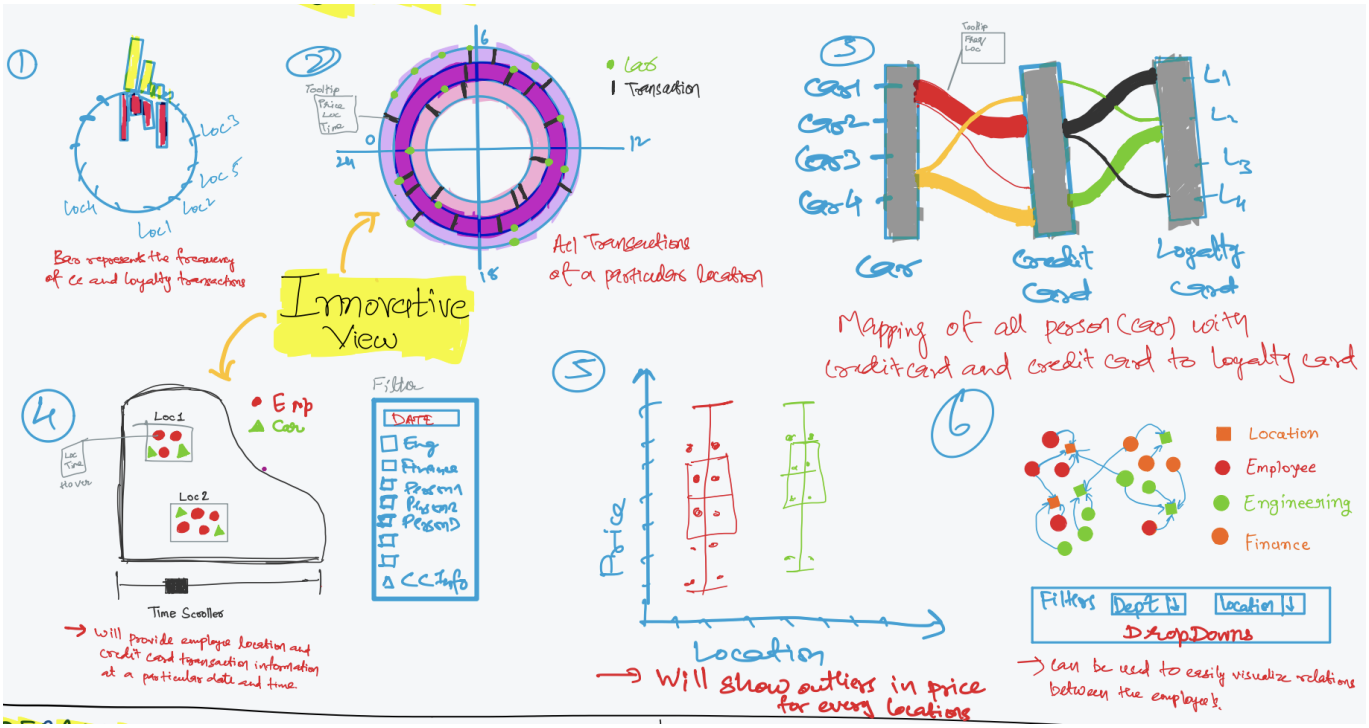
Attribute	Type	Scale
TimeStamp	Ordered,Quantitative,Sequential	13-14 days
ID	Categorical	40
Latitude	Ordered,Quantitative,Sequential	513965
Longitude	Ordered,Quantitative,Sequential	541320

4) Loyalty_data.csv: Tabular

Attribute	Type	Scale
TimeStamp	Ordered,Quantitative,Sequential	13-14 days
Location	Categorical	34
Price	Ordered, Quantitative	1217
Loyalty Number	Categorical	1338

- ESRI shapefiles of Abila and Kronos (in the Geospatial folder)

Mock up image:



Task Abstraction

In order to identify anomalies and suspicious behaviors in the credit card transactions and loyalty card usage data and which people use which credit and loyalty cards, we want to be able to infer certain patterns and oblique information from our visualizations.

Figure 1: Frequency of the transactions of credit card and loyalty card at a location will be shown and users will be able to infer top locations and when they are popular.

Figure 2: User will be able to find the anomalies in the transaction timing.

Figure 3: Credit card and loyalty card matching will be done here and corresponding car owners will be found.

Figure 4: Unofficial and informal relationship among GASTech employees would be observed by analyzing the stop time and traction time of employees.

Figure 5: Anomalies in the price values of the transaction will be inferred using this figure.

Figure 6: Evidence of suspicious relationships between employees will be observed.

Idiom Abstraction

Figure 1:

- This figure will be a Double Circular Bar Chart.

- The locations from the data will be displayed along the circle and for each location we will display 2 bars, the first bar on the outer side of the circle will be for the credit card and the second one on the inner side of the circle will be for the loyalty card.
- The heights of the bars will represent the frequency of the usage or in other words, the number of transactions (either for credit cards or loyalty cards).
- This visualization will help us identify the popularity of the locations.
- When the user hovers on the spokes (or radial bars) of the double circular bar chart, that particular spoke will be highlighted. On clicking that spoke, the multi-circular ring chart will be generated for that location.

Figure 2 (innovative view):

- This visualization will be a Multi Circular Ring Chart, consisting of concentric circles.
- Where every ring will represent a date and the angles will represent the time of the day, for example 0° angle will be 0th hour, 180° angle will be 12th hour and 360° angle will represent the 24th hour of the day.
- The color hue of the rings will be also based on the card usage of that day; the higher the usage, the darker the color.
- ***This chart will be the 1st innovative view of our system. Here, we will be implementing a novel idea of the visualization for analyzing the temporal data.***
- For every transaction on a specific date and time we will create a vertical 5px thick line on a ring which represents that date and at an angle which represents that time of the day.
- Then, we will display the information about when (the time of the day) a car was stationary on each date and we will display this information for all the cars in our database. This information will be displayed by plotting small circles on the respective rings and angles according to the date and time.
- We will provide a checkbox button for this visualization using which the users will be able to add or remove the information about the car's location's (small circles).
- This figure will help us in finding out when a particular location is more famous/visited as well as which cars were present there at that time. Also, we will be able to identify the anomalies in the credit card usage and discrepancies between vehicle, credit, and loyalty card data.
- When the user hovers on a car's channel (dot or star), a tooltip will be displayed with the car's ID and owner information. When the user clicks on the car's channel on the circular rings, the corresponding car and its associated credit cards will be highlighted on the parallel set plot.

Figure 3:

- This will be a Parallel Set Plot consisting of three vertical columns, "CarId", "CreditCardId(last 4 digits)", "LoyaltyCardId".
- For example, a transaction happened at a location where car '1' was present and the credit card used for it was '9999' (we can find it based on the location where the card was used and time of the transaction) then we will create a thin ribbon from CarId=1 to CreditCardId=9999 and similarly we can create a ribbon from CreditCard column to LoyaltyCardId column.
- The width of the ribbons will depend on the number of transactions associated with a particular CarId and CrediCardId and LoyaltyCardId.
- This visualization will help us in identifying the relationship between cars and credit cards as well as loyalty cards which in turn will help us in determining the potential owners of each credit card and loyalty card.
- On clicking a particular car in the cars column, all the other cars and their corresponding links will be grayed out, leaving only the selected car, its linked credit cards and corresponding loyalty cards colored. When the user hovers on the ribbons, a tooltip will display the number of transactions (or in other words the number of times the car was present at a location at which the card was also used in the same time period).

Figure 4 (innovative view):

- This will be a Map Chart showing the geographical top view of the island country of Kronos.
- There will be a range slider under the map using which the users will be able to select the time of the day.
- Then, for that selected time, we will display the information about each employee's location (*here we assume that employee's location is the same as the car's location*) on the map using small circles (each circle will be the representation of an employee).
- And then we will do the circular packing of these circles by grouping them based on how close they are to each other on the map which can be useful in detecting the similarity between the employees.
- ***This chart will be the 2nd innovative view of our system as we will be combining two visualization techniques (map chart with the circular packing) to enhance the user's ability to identify the employee relationships as well as the experience of interacting with the chart.***
- The color of a circle (employee) will be based on the category of the employment to which that employee belongs to.
- There will be a toggle button which can be used to add or remove the information of the credit card transactions. This information will be displayed using small triangles (each triangle will represent a transaction) which will be plotted based on the location of the transaction as well as date and time when it occurred.
- This visualization will be useful for identifying potential informal relationships among GASTech personnel.
- There will be a list of all employees next to the map. Selecting/deselecting any employee from the list will show/hide that employee. By default, we will be showing all the employees. There will also be a checkbox to select/deselect all the employees. There will be a second list next to the employee's list, which will show all the categories. It will have a similar mechanism. There will be a checkbox to show/hide the credit card transactions on the map. There will be a dropdown where the user can select a date.
- The map will show all the locations including the houses of the employees. When the user hovers on the houses, a tooltip will display the information of the employee to whom the house belongs. When the user hovers on the other locations, the tooltip will show aggregated information like the number of people that were present/ had visited that location in the selected time frame (as per the scroller – default time period will be the entire day), the total amount of credit card spending and loyalty card points, etc.
- Selecting (clicking) an employee on the map will generate/change the spider plot.

Figure 5:

- It will be a Boxplot with individual data points (jitter) shown which will be used for analyzing transaction data.
- The X axis will represent the location where a particular transaction occurred.
- The Y axis will represent the amount paid during that transaction.
- Each data point will represent a transaction which will be displayed using small circles.
- This chart will provide statistical information about the credit card usage and some meaningful insights about the distribution of amounts spent at a particular location.
- It will also help us in detecting the outliers in the credit card transaction data such as very large or small amounts spent.
- When the user hovers on the transaction dot, a tooltip will display the name of the person that did the transaction, their credit card no., the date and the time.

Figure 6:

- This figure will be a Network Diagram which will show interconnections between the GAStech employees and the locations they visited during the 14 days.

- The employees will be represented by circular nodes and locations will be represented by square nodes.
- We will create a link between an employee node and location node if that employee has visited that location at any time.
- The colors of the employee nodes will be determined based on their employment category.
- This graph will help us in tracking down the travel patterns of the GAStech employees and it will further improve our assumptions about the relationships between the employees.
- When an employee is selected on figure 4, that employee and his/her connections will be highlighted on this graph.
- Similar to the map, this figure, too, will have a dropdown or a list to select certain departments/employees.

Algorithm Abstraction

Algorithm Abstraction involves creating a set of instructions that can automatically carry out the visual encoding and interaction designs, representing the innermost level of task abstraction.

1. First we have to extract credit card and loyalty card transaction data and calculate transaction frequency for each location.
2. The vehicle data, credit card data and loyalty card data will be merged so that based on the location, amount and time of the transaction, each vehicle could be attached to a credit card and loyalty card, which will help find car owners.
3. The location of each car and its journey throughout the day is tracked based on the geospatial data given to us. Geospatial data has coordinates of the car from which its route will be calculated by calculating the distance between consecutive coordinates given.
4. Cars are classified in two categories: stationary and moving for tracking the route of cars. If the duration between the coordinates is less than 30 sec and the speed is > 20 km/h, cars are classified as moving.
5. If any of the two conditions is violated above, we check the distance the car covers. If it is less than 220m, then cars are classified as stationary or else missing.
6. Time is divided into 5-minute segments, and presence vectors are calculated for cars and credit cards. These presence vectors are checked for similarity. A parallel plot will be plotted showing the similarity between credit cards and vehicles based on the similarity index calculated by presence vectors. Thickness of each line will define the similarity index value.