Simone Pisani – 536959 – Visual Analytics

Project report – VAST Challenge 2021 MC2

Description of data:

The data offered by VAST for this mini challenge consisted of a few files containing the geodata of the city, a png image of the map and fourfour different datasets:

1. car-assignment: {

* LastName
* FirstName
* Id
* CurrentEmploymentType
* CurrentEmploymentTitle

}

1. cc\_data: {

* timestamp 🡪 date + time (hh:mm)
* location
* price
* last4ccnum

}

1. gps: {

* timestamp 🡪 date + time (hh:mm:ss)
* id
* lat
* long

}

1. loyalty\_data: {

* timestamp 🡪 date
* location
* price
* loyaltynum

}

As we can see there were some problems within the data, such as the fact that the different timestamps had different formats or the fact that some fields used as identifiers were not unique, as in the case of the last 4 credit card numbers.

Data preprocessing:

For the type of visualization I had in mind, some data had to be changed.

First of all, I transformed the geospatial data into a geojson format so that I could have used it within d3.

Then I reprocessed the four datasets from a csv format to a json one, mainly for convenience, and then I proceeded to merge the "car-assignment" and "gps" datasets using the "id" field as the key.

The problem here was that the marged file was too large for the upload to Git, so i decided to delete the “CurrentEmplymentType” and “CurrentEmploymentTitle” fields from the file

Design choices:

I decided to divide my project into three sections, each of which contains a visualization and an observation part.

In the first part i wanted to give users an overview of the data, highlighting the amount of transactions for each location in each day using an heatmap. Then i decided to add three card elements containing some statistical informations such as the total number of transactions and the mean, the mode and the standard deviation both from the point of view of places and days.

As for the heatmap, i decided to use it from the early stages of the project, thinking that the choice of using colors to highlight the number of transactions could have been the fastest way to convey the information in a simple and easy-to-read way.

In the second part I choose to create a more focus oriented visualization by providing different information about the transactions of each place on the different days.

This sections consists in a select that allows the user to choose the place and a radio element to define the day. Then these inputs are used as a filter on the original data to create a new dataset to be used in the different charts.

The graphs I originally wanted to use were:

* Piechart: to highlight the percentage of use of both credit and loyalty card, only credit card and only loyalty card in transactions.
* Barchart: to quickly show the amount of transactions on each hour of the day
* Scatterplot: to allow users to view the maximum and minimum price peaks both by day and by hour.

However, I wasn’t satisfied with the result: I wanted to add some more specific information showing the relationship between the transaction, the price and the time. After some thoughts i decided to use a Sankey diagram to get a complete view of the links between credit card, loyalty card, price and time of the transaction.

In the third and final part I decided to analyze geospatial data, creating a geojson dataset from gps data.

In this section I created an svg element in which I then inserted the map of the island, overlapping the coordinates of the streets on it and then making their color transparent.

Subsequently, starting from the two main filters which are a radio element to choose the day and a range element for the time, I decided to display the current position of each person on the map, adding a small trail to highlight recent movements .

Then I added a command line, to automate the progression or regression of the timeline, and two card elements where are shown, in a dynamic way, the list of transactions in the last 30 minutes and the people's movements (given that the data is not always continue I entered an error of plus or minus 5 seconds).

In the left part of the section I then added a legend containing a list of the different inhabitants of the island, with the possibility of hiding or showing them on the map by clicking on the element

regarding the chromatic choice, I decided to use blue and gray and their different scales, mainly because the contrast was strong enough to allow them to be distinguished correctly and because I thought their combination would not tire the eyes too much .

Regarding the piechart I chose three very different colors such as ivory, green and red, while for the scatterplot I preferred to use blue and red but adopting soft shades to not distract too much from the main chart of the section which is the sankey diagram

In conclusion, for the map in the third section, i decided to use black to highlight the people because it’s a color that is easy to distinguish and it didn’t standoff too much by overlaying it on the map.

State-of-art:

As for the graphs, I chose to use them almost immediately by scrolling through the d3 documentation and finding some particularly suited to my visualization idea. The one exception was the sankey chart which I chose after running through several linked chart alternatives.

As for the map, I wanted to use it from the early stages of the project while the idea of adding the traces of the previous movements came to me while observing one of the graphs during one of the last lessons of the course.

Use case examples:

this part has already been inserted within the project in the observation sections.

The points of the challenge that I have tried to solve and answer are:

* Using just the credit and loyalty card data, identify the most popular locations, and when they are popular. What anomalies do you see?
* What discrepancies between vehicle, credit, and loyalty card data do you find?
* Can you infer the owners of each credit card and loyalty card?
* Given the data sources provided, identify potential informal or unofficial relationships among GASTech personnel. Provide evidence for these relationships.
* Do you see evidence of suspicious activity?