**CSE564 Final Project Prelim Report**

Xinhe Huang 108390641

Silu Ou 110563639

**Problem & background**

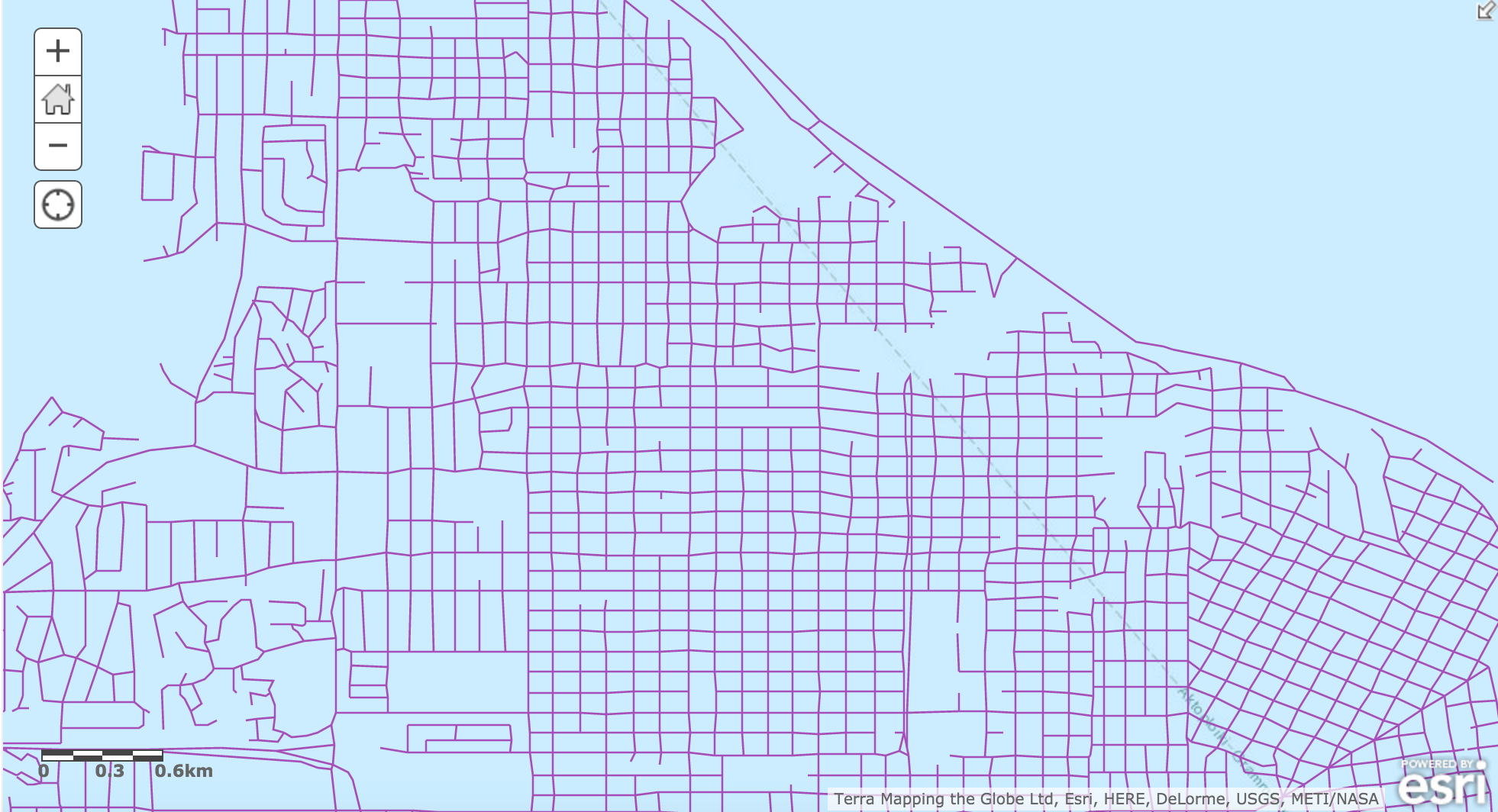
Some employees of GAStech disappeared in the last two weeks. Since they used the cars of the company and the data of these cars’ movements as well as the data of the employees’ transactions were both recorded, we can analyze these data to find out the reason of their disappearance.

In order to find out why they disappeared, we need to answer the following questions:

1. What does a day in the life of a typical GAStech employee look like?
2. What unusual events or patterns do we observe? For this question we need to find out the location and the person involved as well as the time when an event occured. We also need to decide if it is necessary to pay attention to this pattern.
3. How to deal with the imperfect data?

**Preparation**

In order to get a better understanding in the *gps.csv* file, we use the app ArcGIS to open the ESRI shapefile *Abila* and translate the map into a file named *location.csv*. This file records the gps of all the locations shown in the map. Then we can translate the records of latitude and longitude in the *gps.csv* file into corresponding locations. The ESRI shapefile *Abila* and the map are shown as below.



Abila



map

We compare these two files carefully and record the range of each location’s gps. Then we use the average value of the range as the gps of one location.

**Data Analysis**

First we take a look at each csv file roughly. We find that *gps.csv* file records each car’s gps at every second as long as the car is moving. Thus we need to handle the data in the *gps.csv* file to get the start location and end location of each journey. Then for each car we calculate the frequency of its visit to each place. If one place has a frequency far below others, it means this car seldom visits this place, which is an unusual behavior. We also need to calculate the time period of each car’s journey since for the user of one car he/she should have a similar schedule for each day. If one day there is a journey having a time period that is far from other days, it means that on this day the user has an unusual behavior.

Since we haven’t finished the visualization of this part, the interaction of the data in the back end is shown below.



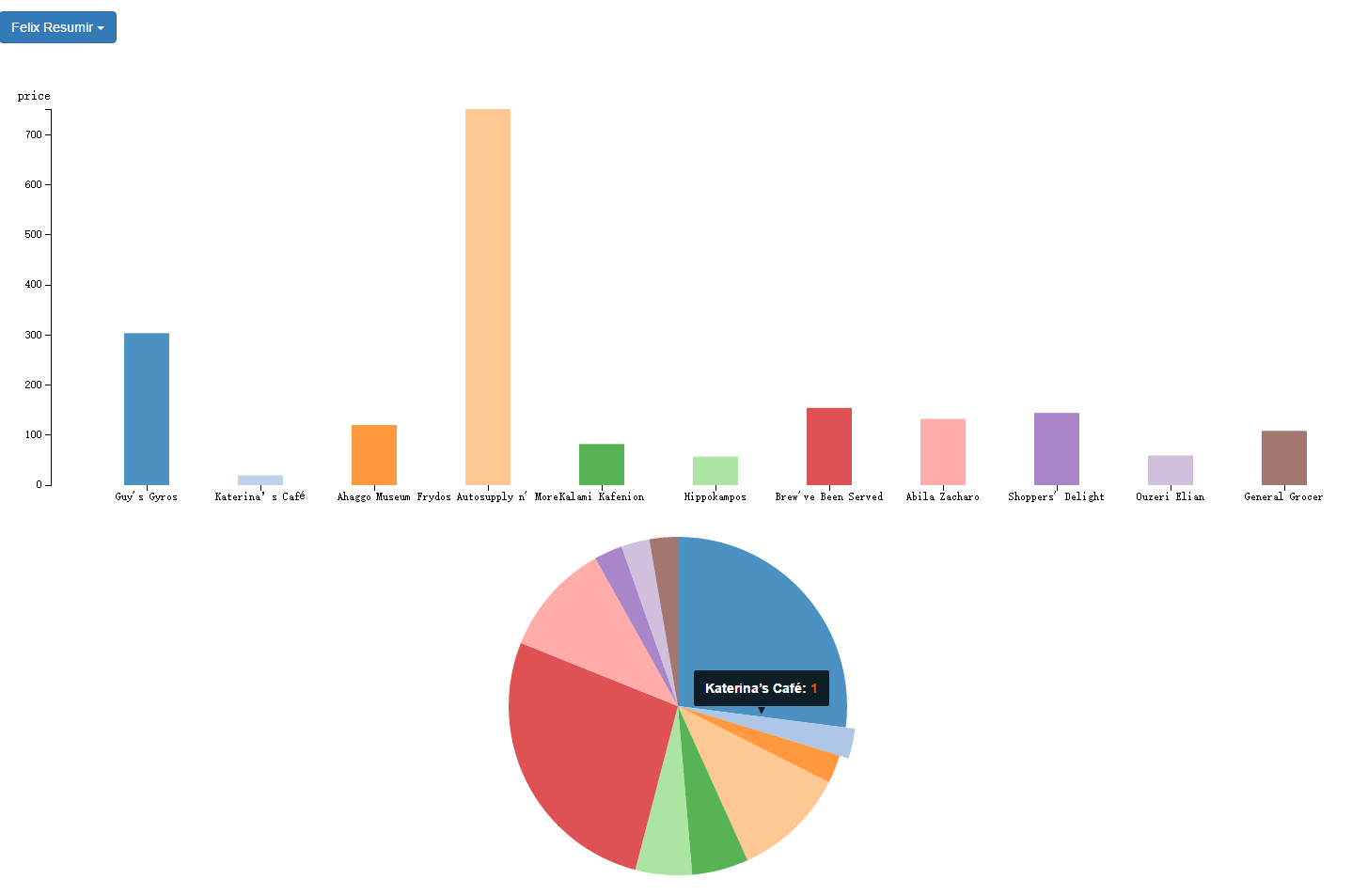
The picture shows the data of the start locations and the end locations of one car with id 35. We obtain these data from the original data in *gps.csv* file.

For *cc\_data.csv* file we do the similar analysis work. But before we analyze the data, we first need to filter them. We find that some people only go to one or two places and have tractions there. It is not necessary to do analysis on these people since we can’t get a lot of useful information from their monotonous life. Thus we need to delete these people’s records from the original data. Then we calculate the frequency of each person’s transaction at each location to find out their unusual behavior. In addition, we calculate the total price of each person’s transaction at each location. Then we compare the total price with the frequency of its corresponding location to see if there is any abnormality.

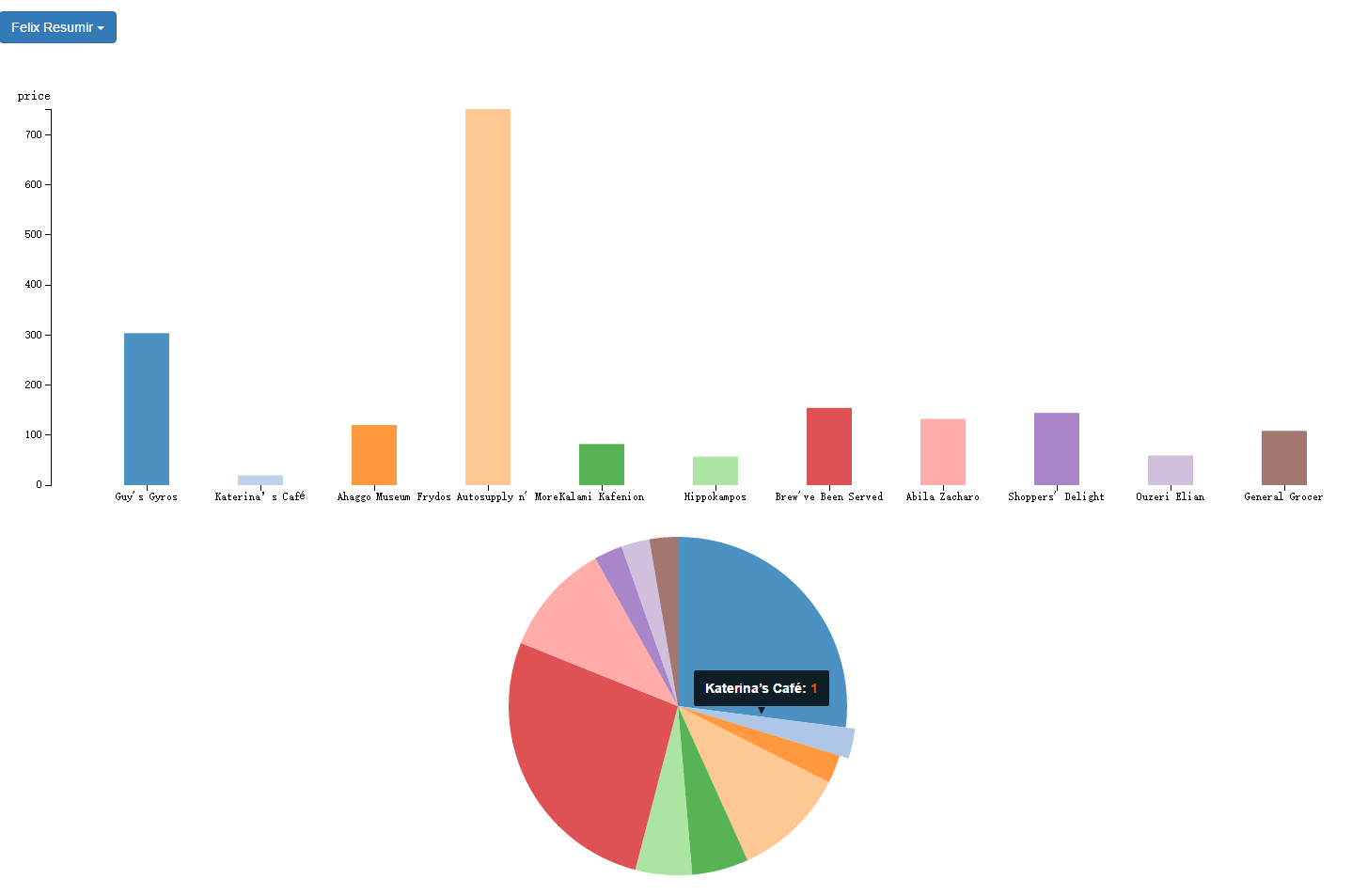
**Charts/Graphs**

In this section we will give some examples that we do analysis on. We choose 3 people to show how we find their unusual behaviors.

1.Felix Resumir



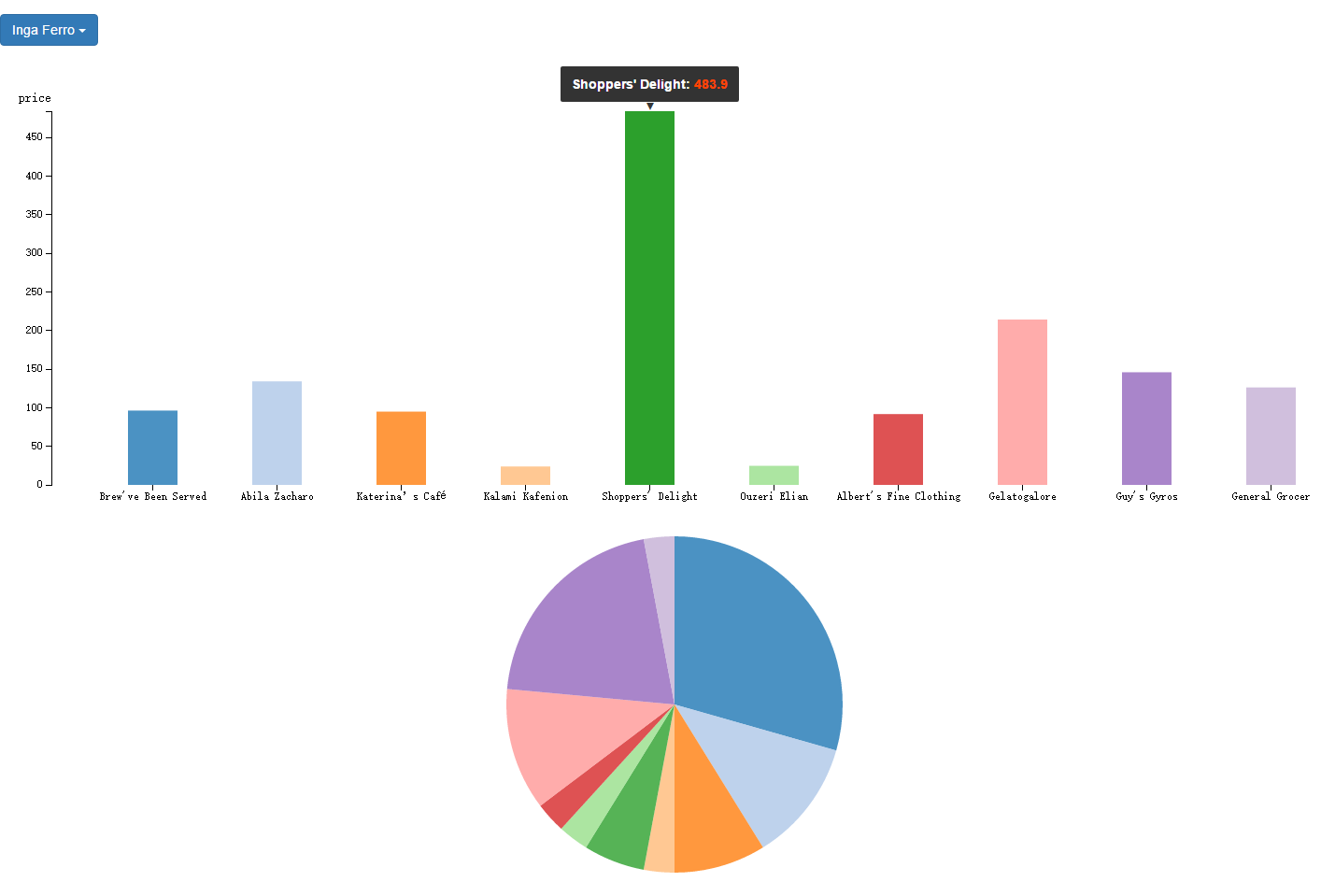
bar chart



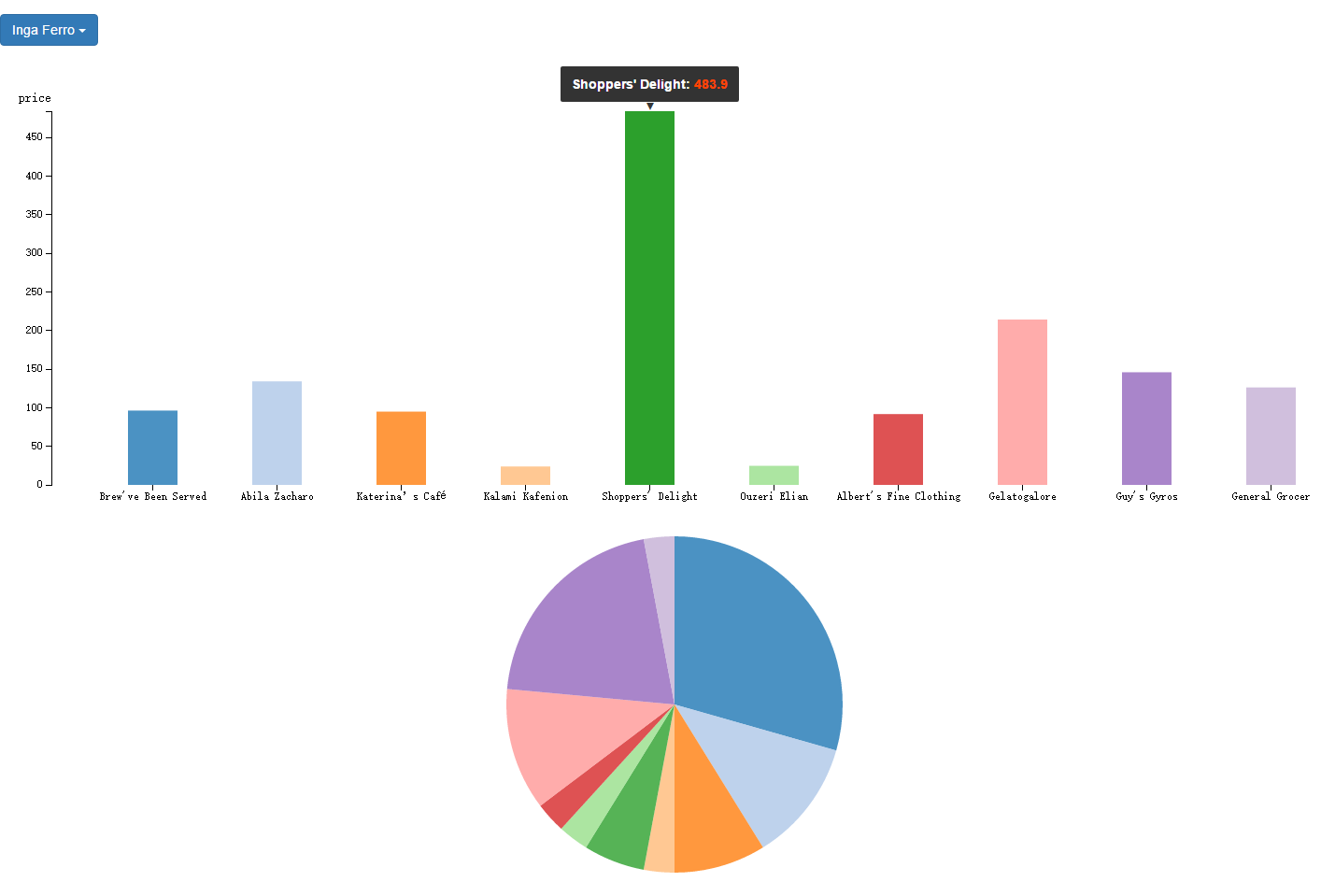
pie chart

From the charts above, we know that Felix just made one transaction at Katerina’s Café while he made more transactions at other locations during two weeks. This transaction is an unusual behavior which needs our analysis.

2.Inga Ferro



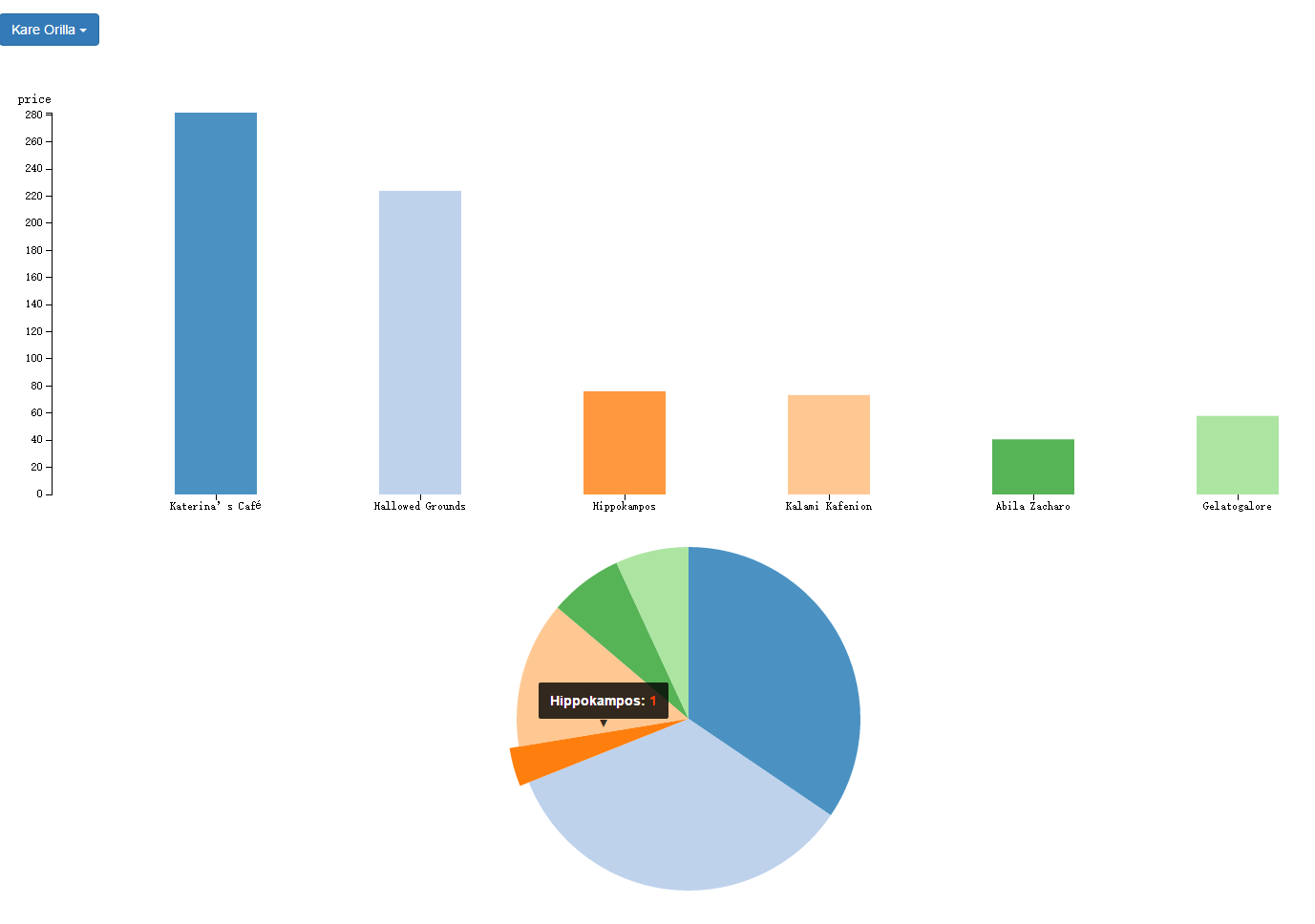
bar chart



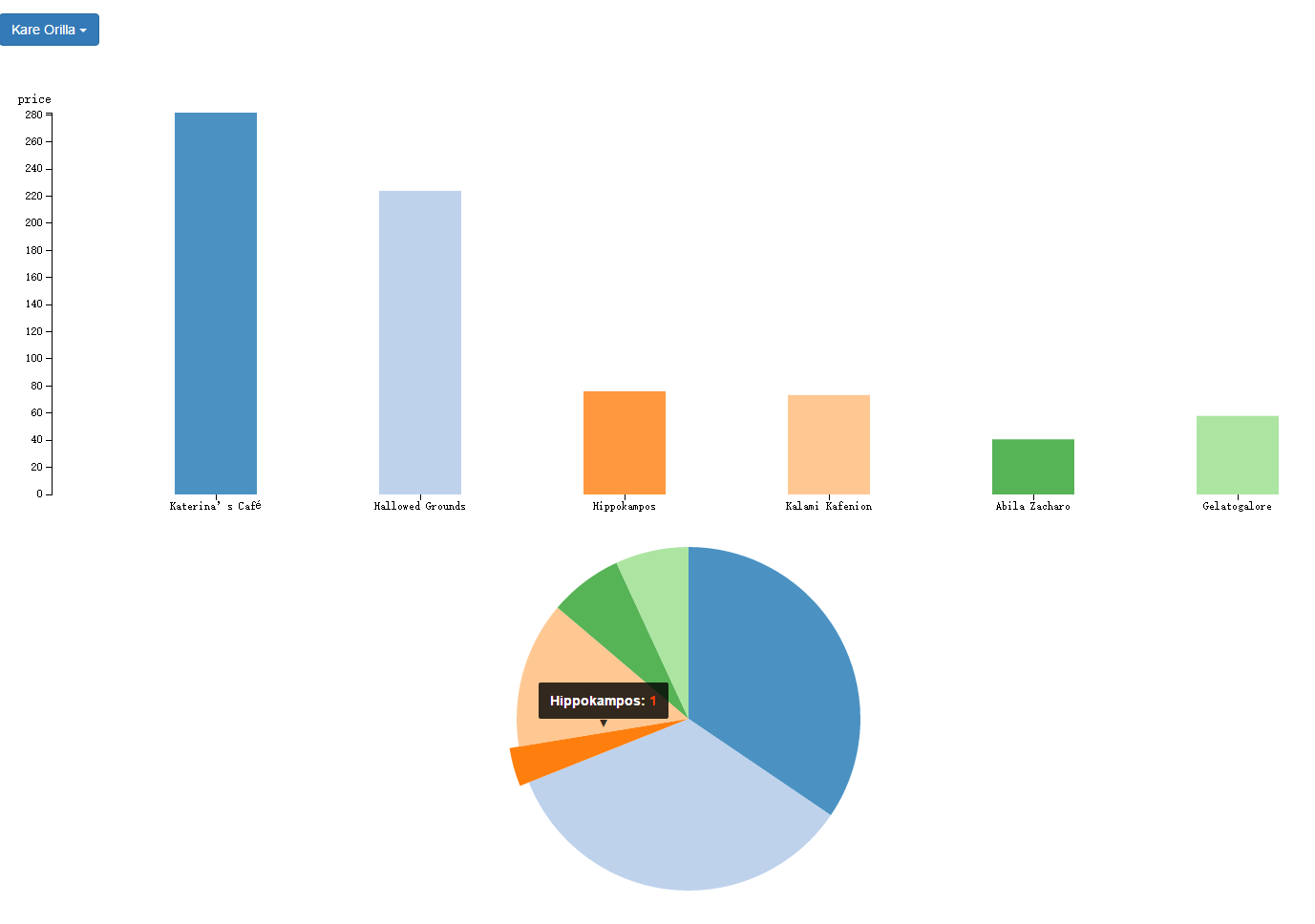
pie chart

This guy, Inga Ferro, shopped twice at Shoppers’ Delight during two weeks, however, he spent most ($483.9) on it. We will pay more attention to this behavior to see if there is any abnormality.

3. Kare Orilla



bar chart



pie chart

Compared to others, this person has a regular life. He went to Katerina’s Cafe and Hallowed Grounds almost every day and sometimes he went to other stores except “Hippokampo”. Since he only went to this place once during the last two weeks, this behavior will be analyzed.

**Future work**

Since we have spent a lot of time on the preparation (translate the map into *location.csv* file), we need to finish the visualization of our analyzed data in the next step. We also need to do further analysis in those unusual behaviors we have noticed in order to find out the relations between these abnormalities and the employees’ disappearance.

Thus in the future, we are going to make specific graphs according to the gps data in *gps.csv* file and *location.csv* file. Also we will compare the locations between these gps data and the location data in *cc\_data.csv* file. Then we can get the answers to the questions.