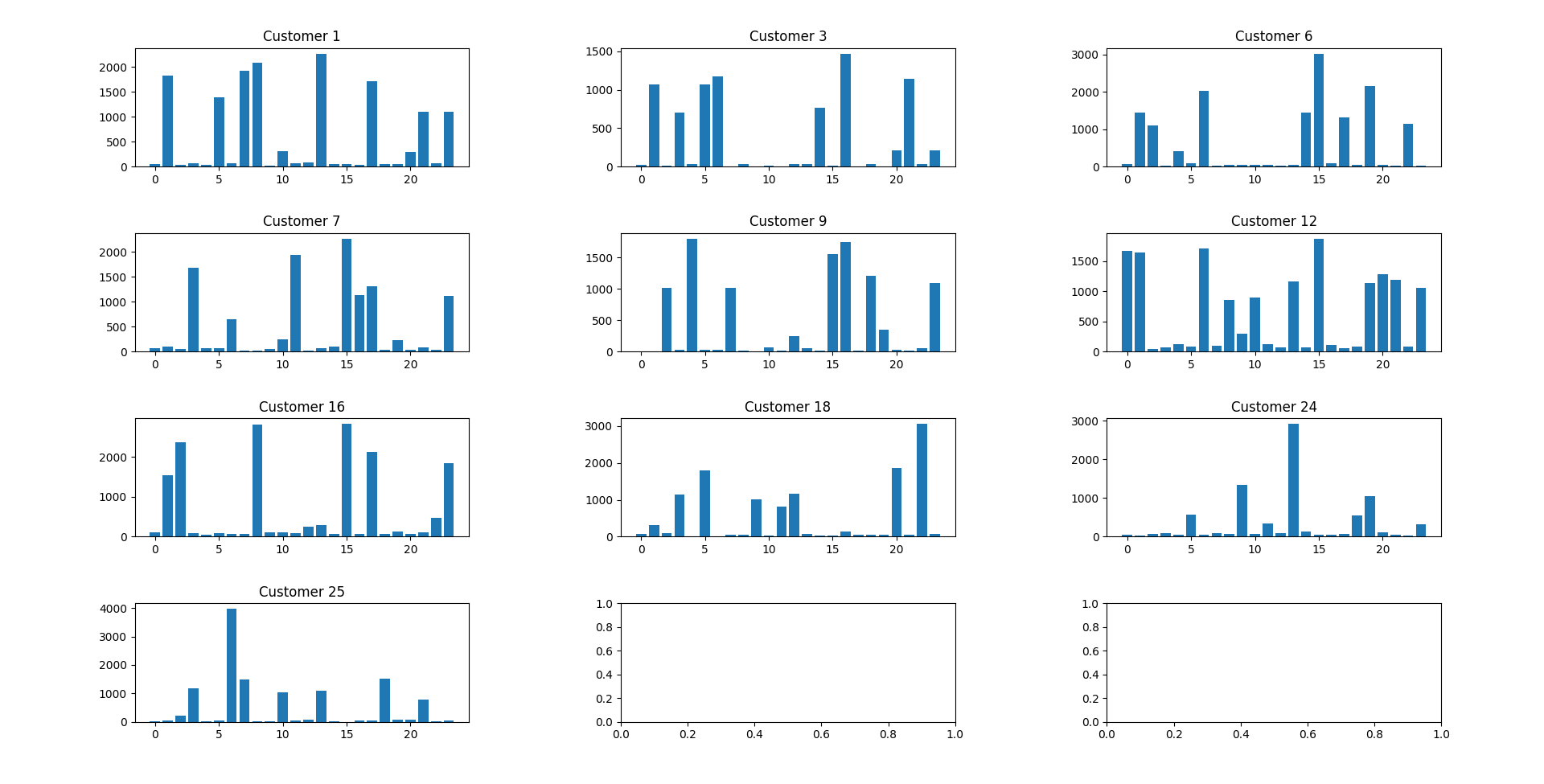
**Project : Fraud Detection**

Source: Files from Maitree7 on Github.

Pre-process:

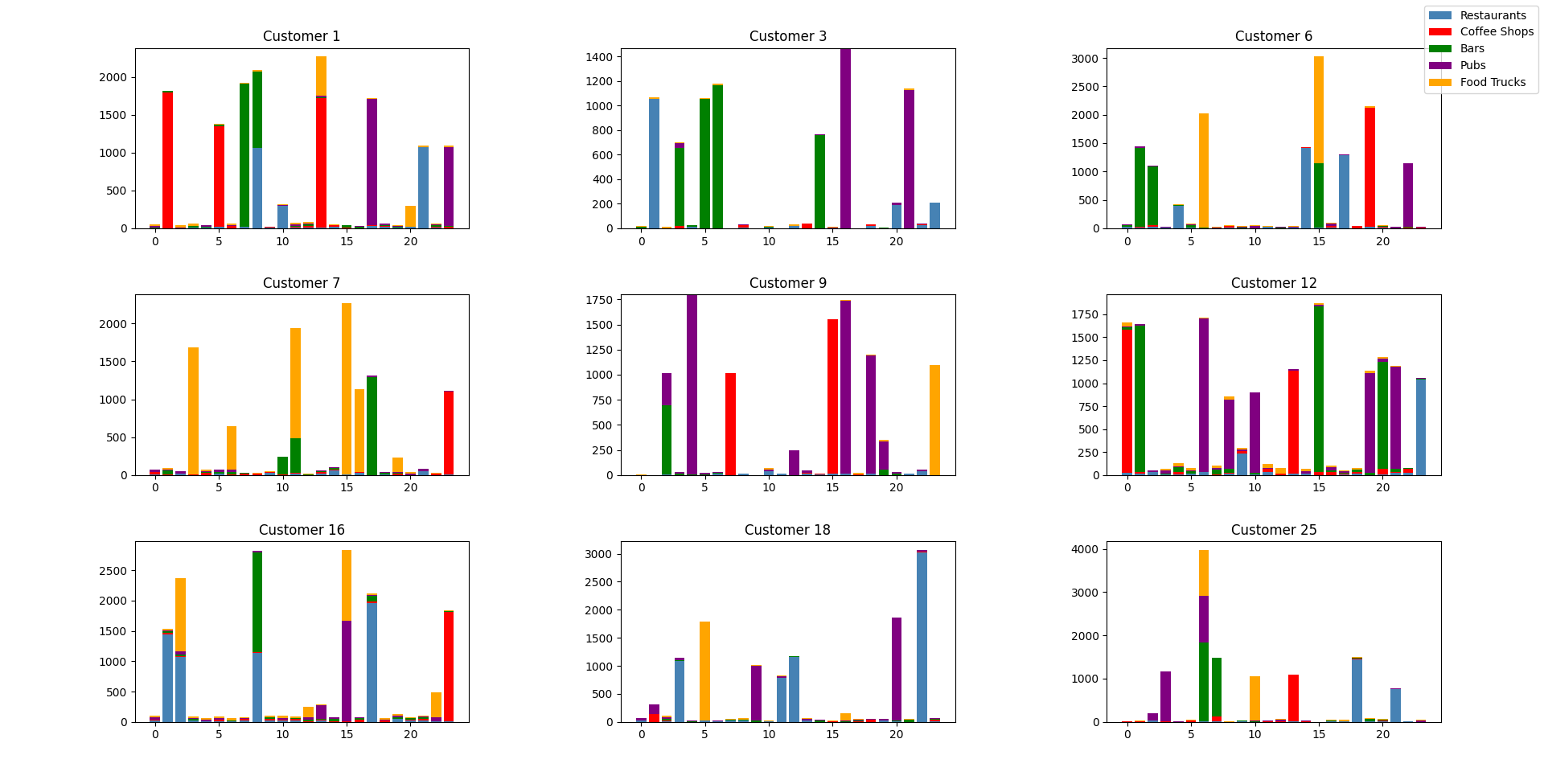
* Using SQL, read from csv into tables and join all tables into a comprehensive table of all transactions.
* Using SQL to better understand certain trends in the transactions.

After determining which customers had multiple transaction days over $1000, I produced the following bar graph to document the times in which most of their transactions occurred.



All customers except #24 have a large amount of money that was spent at an odd time (whether it is 2am or 11pm).

Grouping the data and removing customer #24 from the plot, I produced this color-coded plot.





You can tell that there is some obvious fraud going on. It is very unlikely that a person would spend over a thousand dollars at a coffee shop at 1am (Customer 1 and 12) or $1500 at a food truck at 3am (Customer 7). Especially since these customers don’t appear to spend this much during other parts of the day. Does customer 18 simply always get breakfast at a food truck or are they a victim of credit card fraud. The blue arrow shows some obvious areas where potential fraud could be happening.

|  |  |  |
| --- | --- | --- |
| Customer | Customer 3 | Customer 6 |
| Customer 7 | Customer 9 | Customer 12 |
| Customer 16 | Customer 18 | Customer 25 |

Highlighted are potential areas where there may be fraud. I determined these using the following logical criteria:

1. Places where large amounts were spent at locations that are not normally frequented during a particular hour, for instance a coffee shop is usually open from 6am-2pm, thus having large transactions at 1am, 11pm etc. is out of the ordinary. Frequenting Bars at 5am is another unlikely situation.
2. Places where large amounts were spent, but the hour prior and after had minimal spending. This suggests that the large amount was spent on a lump sum of transactions and was likely not the customer.

Note: Even though an area was highlighted, it does not mean that every transaction in this category and hour was indeed fraud. These tables depict the **sum** of transactions for the duration of documented transactions within the given time and merchant type.

Next Step: Look at these key points more in depth to discover the actual instances of fraud. If this were a bank, the protocol would be to verify key transactions with the customer, dispute charges and reorder their card.