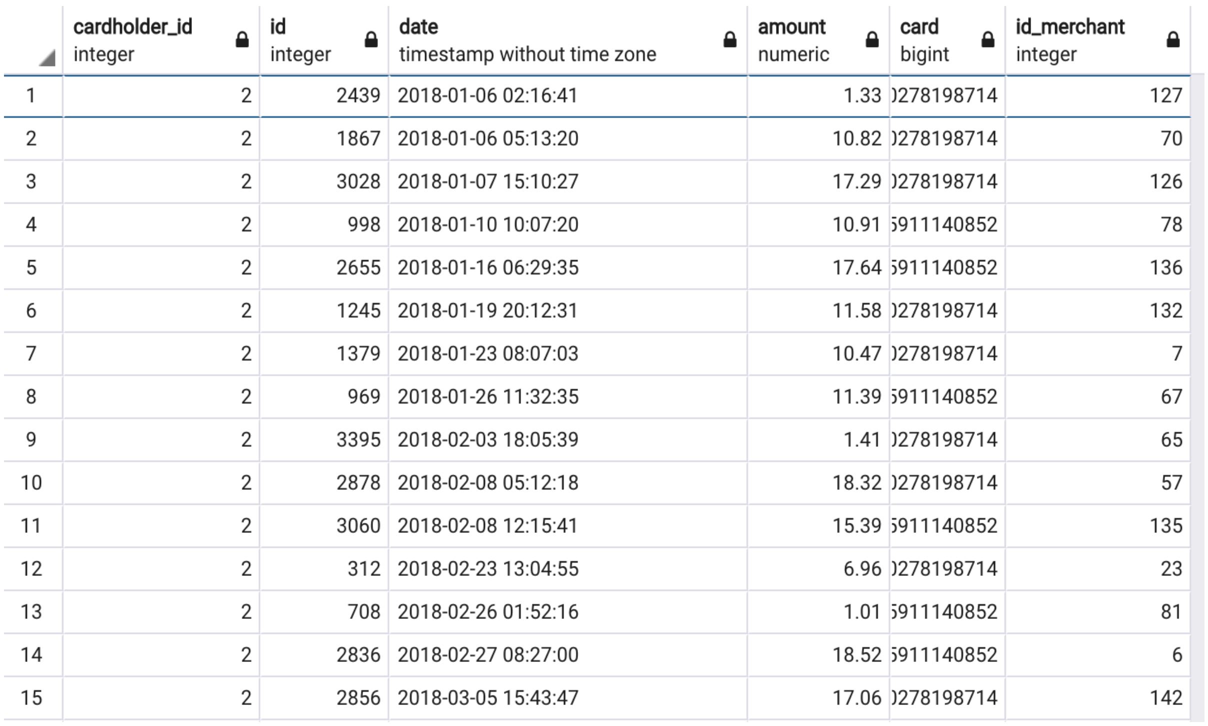
Final Report – Fraud Transactions Analysis

1. Transactions per Card Holder:

a. The following script isolates the transactions for n specific Card Holder (as an example Card Holder with id

```
SELECT cc.cardholder_id, t.id, t.date, t.amount, t.card, t.id_merchant
FROM transaction t JOIN credit_card cc
ON (t.card = cc.card) AND (cc.cardholder_id = 2)
ORDER BY cc.cardholder_id, t.date;
```

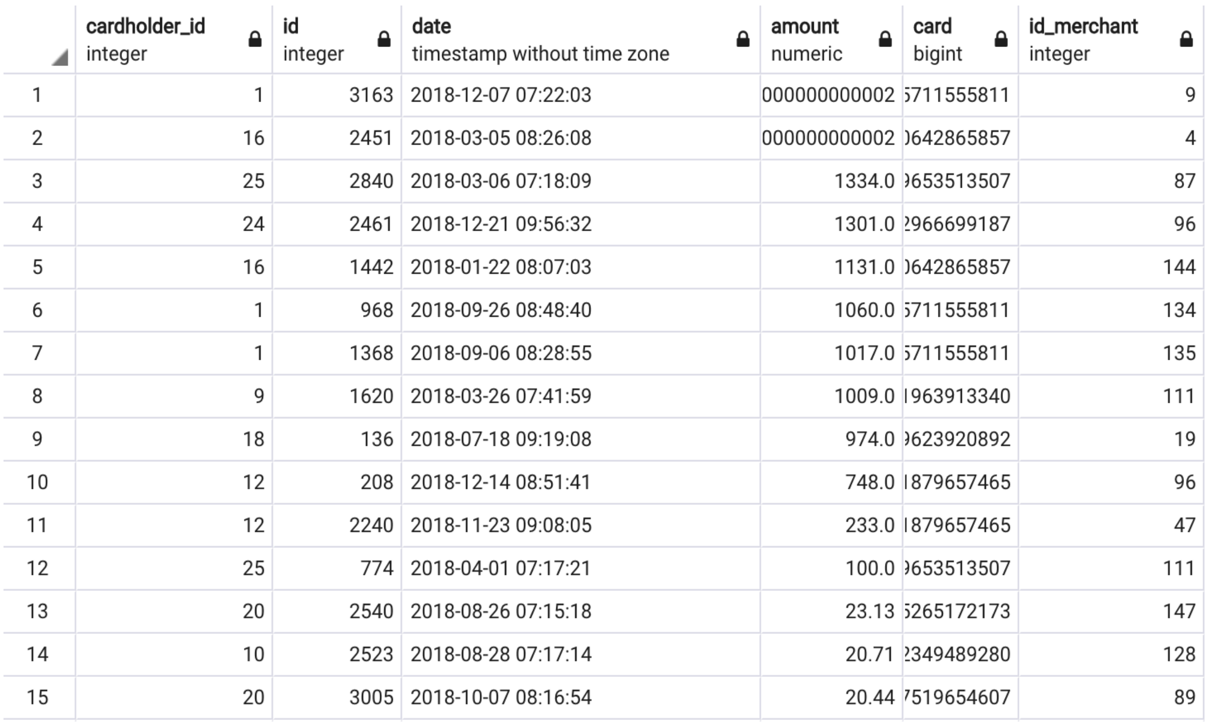


	cardholder_id integer	id integer	date timestamp without time zone	amount numeric	card bigint	id_merchant integer
1	2	2439	2018-01-06 02:16:41	1.33	0278198714	127
2	2	1867	2018-01-06 05:13:20	10.82	0278198714	70
3	2	3028	2018-01-07 15:10:27	17.29	0278198714	126
4	2	998	2018-01-10 10:07:20	10.91	5911140852	78
5	2	2655	2018-01-16 06:29:35	17.64	5911140852	136
6	2	1245	2018-01-19 20:12:31	11.58	0278198714	132
7	2	1379	2018-01-23 08:07:03	10.47	0278198714	7
8	2	969	2018-01-26 11:32:35	11.39	5911140852	67
9	2	3395	2018-02-03 18:05:39	1.41	0278198714	65
10	2	2878	2018-02-08 05:12:18	18.32	0278198714	57
11	2	3060	2018-02-08 12:15:41	15.39	5911140852	135
12	2	312	2018-02-23 13:04:55	6.96	0278198714	23
13	2	708	2018-02-26 01:52:16	1.01	5911140852	81
14	2	2836	2018-02-27 08:27:00	18.52	5911140852	6
15	2	2856	2018-03-05 15:43:47	17.06	0278198714	142

2. Consider the time period 7:00 a.m. to 9:00 a.m.

a. What are the top 100 highest transactions during this time period?

```
SELECT cc.cardholder_id, t.id, t.date, t.amount, t.card, t.id_merchant
FROM transaction t JOIN credit_card cc
ON (t.card = cc.card)
WHERE (extract(hour from t.date) >= 7) AND (extract(hour from t.date) <= 9)
ORDER BY t.amount DESC LIMIT 100;
```



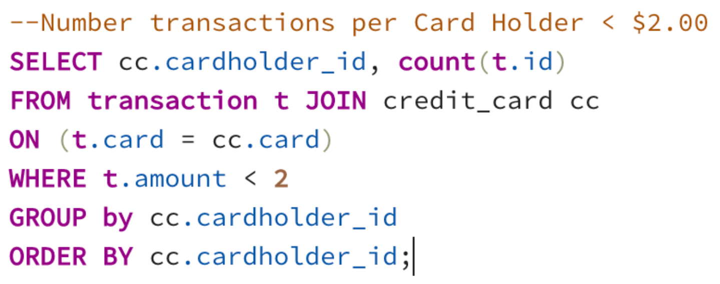
	cardholder_id integer	id integer	date timestamp without time zone	amount numeric	card bigint	id_merchant integer
1	1	3163	2018-12-07 07:22:03	000000000002	5711555811	9
2	16	2451	2018-03-05 08:26:08	000000000002	0642865857	4
3	25	2840	2018-03-06 07:18:09	1334.0	9653513507	87
4	24	2461	2018-12-21 09:56:32	1301.0	2966699187	96
5	16	1442	2018-01-22 08:07:03	1131.0	0642865857	144
6	1	968	2018-09-26 08:48:40	1060.0	5711555811	134
7	1	1368	2018-09-06 08:28:55	1017.0	5711555811	135
8	9	1620	2018-03-26 07:41:59	1009.0	1963913340	111
9	18	136	2018-07-18 09:19:08	974.0	9623920892	19
10	12	208	2018-12-14 08:51:41	748.0	1879657465	96
11	12	2240	2018-11-23 09:08:05	233.0	1879657465	47
12	25	774	2018-04-01 07:17:21	100.0	9653513507	111
13	20	2540	2018-08-26 07:15:18	23.13	5265172173	147
14	10	2523	2018-08-28 07:17:14	20.71	2349489280	128
15	20	3005	2018-10-07 08:16:54	20.44	7519654607	89

b. Do you see any fraudulent or anomalous transactions?

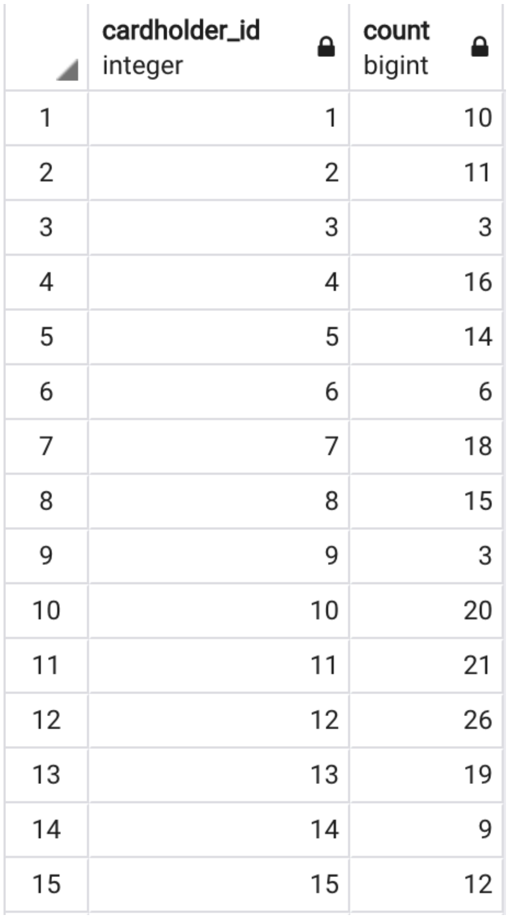
Not really, because we should compare against the trends per Card Holder. Even when we can see some very big transactions, we cannot infer that these ones are frauds.

3. Some fraudsters hack a credit card by making several small payments (generally less than $2.00), which are typically ignored by cardholders.

a. Count the transactions that are less than $2.00 per cardholder. Is there any evidence to suggest that a credit card has been hacked? Explain your rationale.



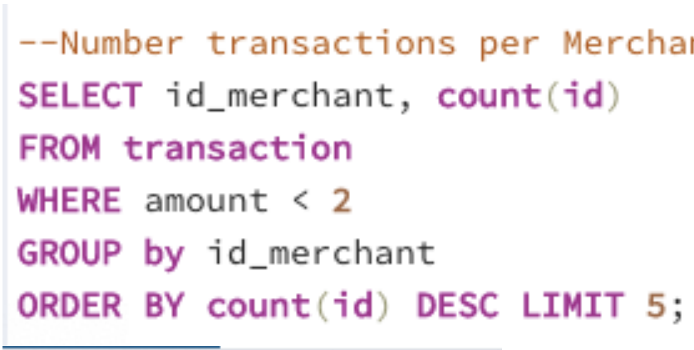
```
--Number transactions per Card Holder < $2.00
SELECT cc.cardholder_id, count(t.id)
FROM transaction t JOIN credit_card cc
ON (t.card = cc.card)
WHERE t.amount < 2
GROUP by cc.cardholder_id
ORDER BY cc.cardholder_id;
```



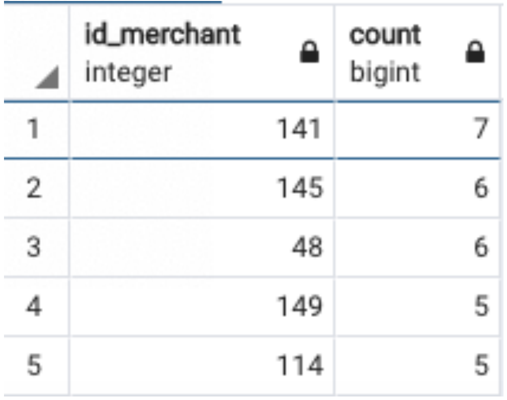
	cardholder_id integer	count bigint
1	1	10
2	2	11
3	3	3
4	4	16
5	5	14
6	6	6
7	7	18
8	8	15
9	9	3
10	10	20
11	11	21
12	12	26
13	13	19
14	14	9
15	15	12

Even when we can see a number of transactions < $2.00, it doesn't mean that these are signals of fraud. Again, we should analyze case by case to see what is the behavior per Card Holder.

4. **What are the top 5 merchants prone to being hacked using small transactions?**

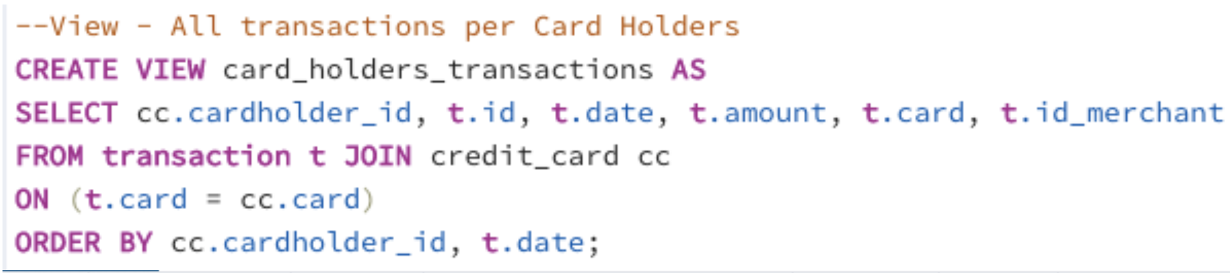


```
--Number transactions per Merchant < $2.00 - TOP 5
SELECT id_merchant, count(id)
FROM transaction
WHERE amount < 2
GROUP by id_merchant
ORDER BY count(id) DESC LIMIT 5;
```

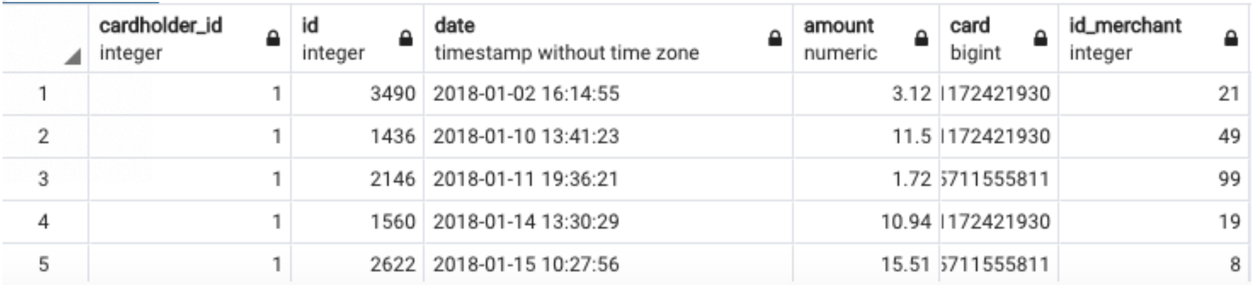


	id_merchant integer	count bigint
1	141	7
2	145	6
3	48	6
4	149	5
5	114	5

5. **Once you have a query that can be reused, create a view for each of the previous queries.**



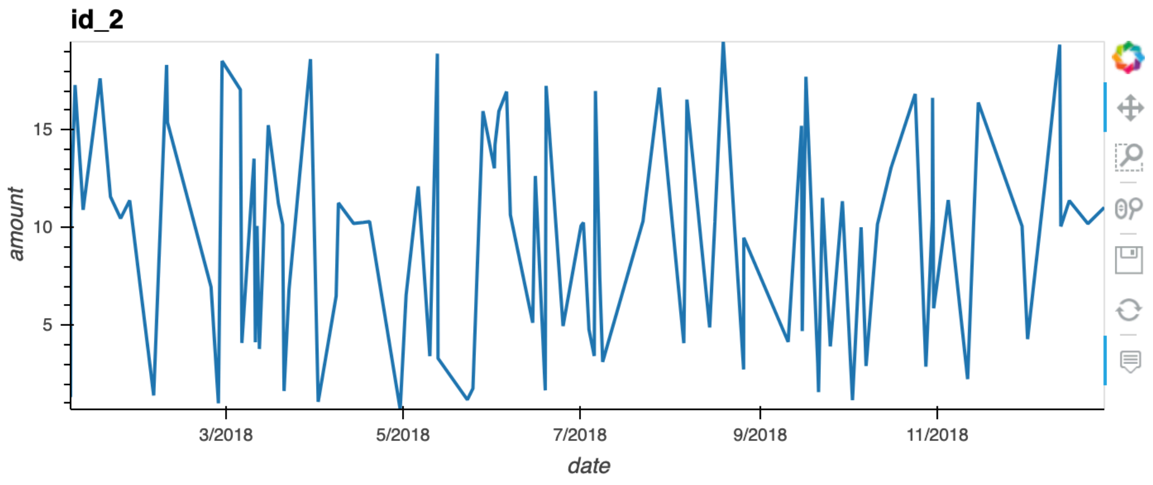
```
--View - All transactions per Card Holders
CREATE VIEW card_holders_transactions AS
SELECT cc.cardholder_id, t.id, t.date, t.amount, t.card, t.id_merchant
FROM transaction t JOIN credit_card cc
ON (t.card = cc.card)
ORDER BY cc.cardholder_id, t.date;
```

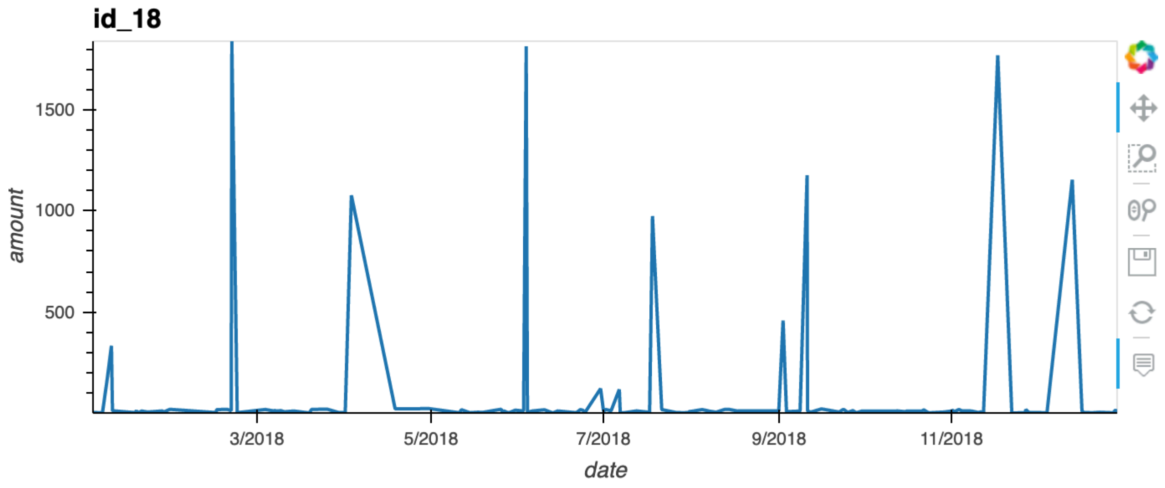


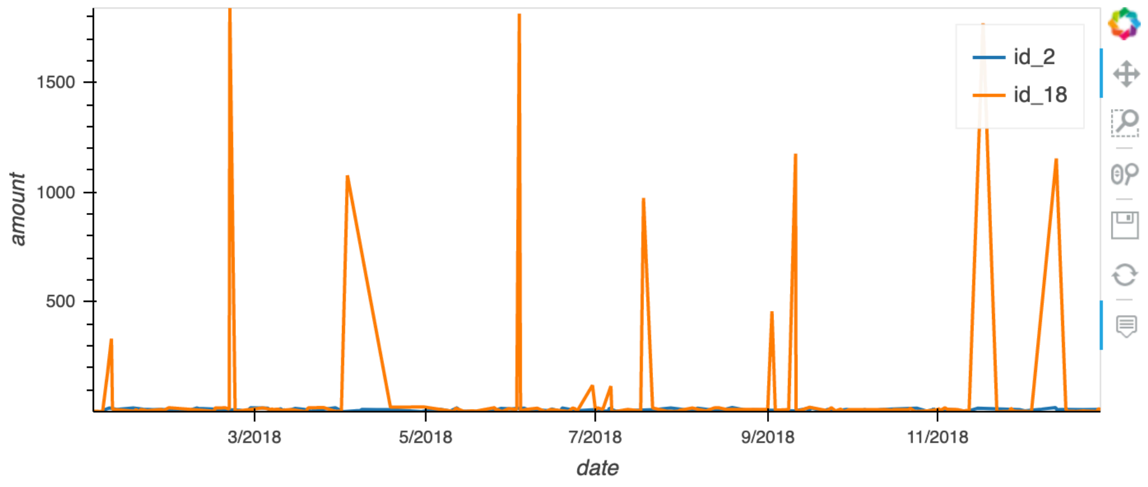
	cardholder_id integer	id integer	date timestamp without time zone	amount numeric	card bigint	id_merchant integer
1	1	3490	2018-01-02 16:14:55	3.12	172421930	21
2	1	1436	2018-01-10 13:41:23	11.5	172421930	49
3	1	2146	2018-01-11 19:36:21	1.72	711555811	99
4	1	1560	2018-01-14 13:30:29	10.94	172421930	19
5	1	2622	2018-01-15 10:27:56	15.51	711555811	8

6. Report for Fraudulent Transactions:

a. Verify if there are any fraudulent transactions in the history of two of the most important customers of the firm. For privacy reasons, you only know that their cardholders' IDs are 18 and 2.





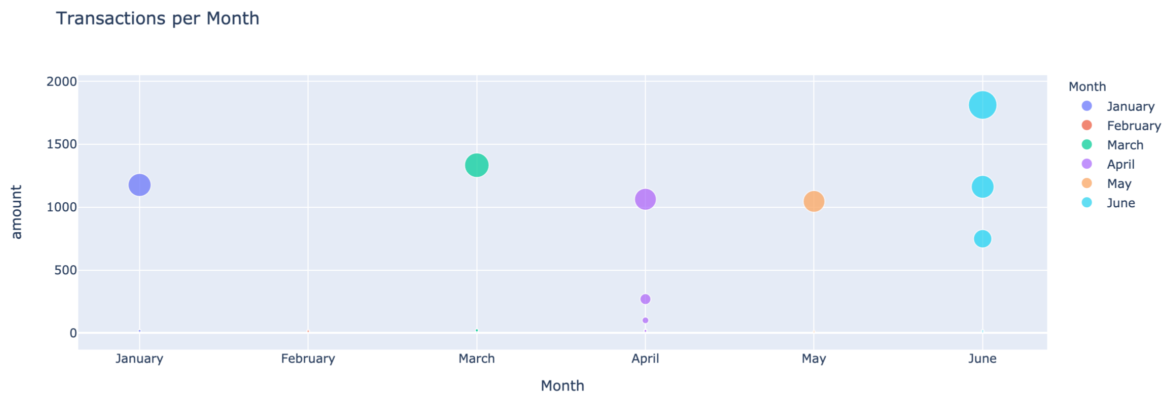


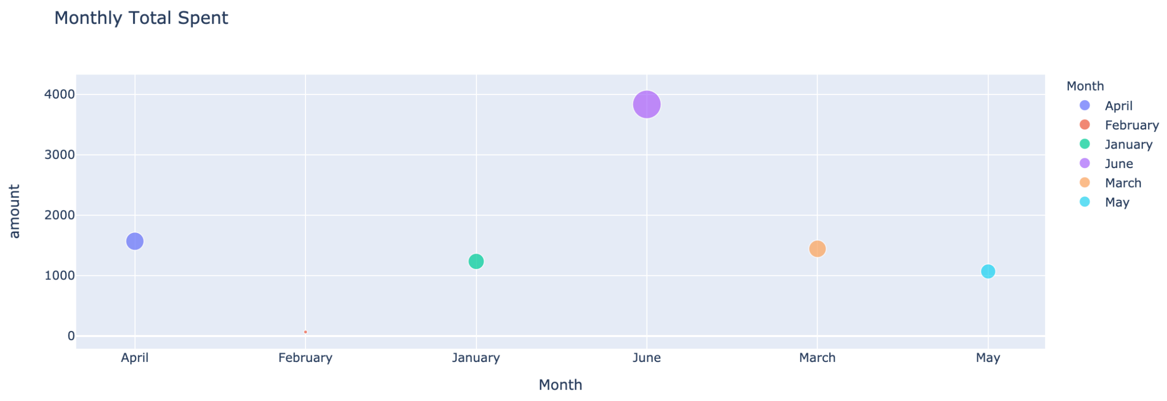
I think both Card Holders with ids 2 and 18 have a pretty stable consumption patterns for the most of the dates with small amounts most of the time, but in some dates, we can see that Card Holder with id 18 has really higher charges. This situation for Card Holder id 18 should be investigated to realize if those are really signals of frauds or simply that Card Holder has had extraordinary spends during the observed dates.

b. The CEO of the biggest customer of the firm suspects that someone has used her corporate credit card without authorization in the first quarter of 2018 to pay quite expensive restaurant bills. You are asked to find any anomalous transactions during that period.

   Using Plotly Express, create a series of six box plots, one for each month, in order to identify how many outliers per month for cardholder ID 25.

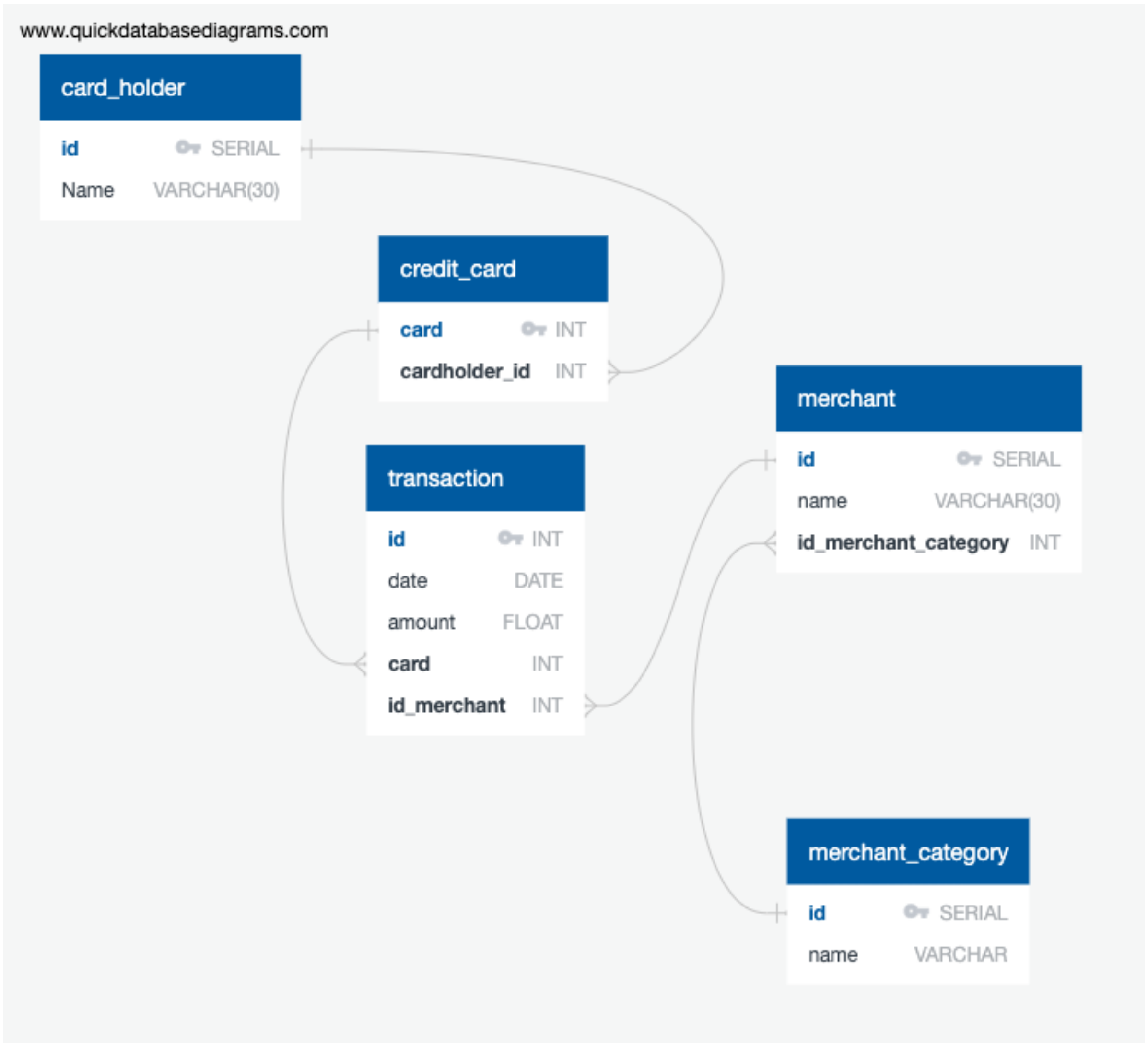
   Do you notice any anomalies? Describe your observations and conclusions.





As we can see, there are some high charges for Card Holder id 25 during the six months, but especially in June we can see a much higher spend. That could be a signal of fraud, thus it should be investigated and checked with the Card Holder/CEO.

7. DB ERD:



www.quickdatabasediagrams.com
card_holder
id SERIAL
Name VARCHAR(30)
credit_card
card INT
cardholder_id INT
transaction
id INT
date DATE
amount FLOAT
card INT
id_merchant INT
merchant
id SERIAL
name VARCHAR(30)
id_merchant_category INT
merchant_category
id SERIAL
name VARCHAR