In [1]:
```
# Initial imports
```

In [2]:
```
import pandas as pd
import calendar
import plotly.express as px
import hvplot.pandas
from sqlalchemy import create_engine
```

In [ ]:
```
```

In [3]:
```
# Create a connection to the database for all 5 tables
```

In [4]:
```
##from sqlalchemy import create_engine
engine = create_engine("postgresql+psycopg2://postgres:Iwantcash2274@localhost:5432/wee
```

In [5]:
```
query1 = 'select * from card_holder;'
query2 = 'select * from credit_card;'
query3 = 'select * from data_merchant;'
query4 = 'select * from merchant_category;'
query5 = 'select * from data_transaction;'
```

In [6]:
```
card_holder = pd.read_sql(query1, engine)
card_holder.head()
```

Out[6]:

	id	name
0	1	Robert Johnson
1	2	Shane Shaffer
2	3	Elizabeth Sawyer
3	4	Danielle Green
4	5	Sara Cooper

In [7]:
```
credit_card = pd.read_sql(query2, engine)
credit_card.head()
```

Out[7]:

	card	cardholder_id
0	3517111172421930	1
1	4761049645711555811	1
2	4866761290278198714	2
3	675911140852	2
4	30078299053512	3

In [8]:
```
data_merchant = pd.read_sql(query3, engine)
data_merchant.head()
```

Out[8]:

	id	name	id_merchant_category
0	1	Murphy, Heath and Fields	1
1	2	Riggs-Adams	1
2	3	Sanders, Parks and Mcfarland	2
3	4	Mccarty-Thomas	3
4	5	Miller-Blevins	4

In [9]:
```
merchant_category = pd.read_sql(query4, engine)
merchant_category.head()
```

Out[9]:

	id	name
0	1	restaurant
1	2	coffee shop
2	3	bar
3	4	pub
4	5	food truck

In [10]:
```
data_transaction = pd.read_sql(query5, engine)
data_transaction.head()
```

Out[10]:

	id	date	amount	card	id_merchant
0	222	2018-01-01 21:35:10	6.22	3561954487988605	69
1	2045	2018-01-01 21:43:12	3.83	5135837688671496	85
2	395	2018-01-01 22:41:21	9.61	213193946980303	82
3	3309	2018-01-01 23:13:30	19.03	4263694062533017	5
4	567	2018-01-01 23:15:10	2.95	4498002758300	64

In [ ]:
```

```

In [11]:
```
# Data Analysis Question 1
```

In [12]:
```
# combine all tables
query_fraud = """
select card_holder.id,
data_transaction.date, data_transaction.amount from data_transaction
join credit_card on credit_card.card = data_transaction.card
join card_holder on card_holder.id = credit_card.cardholder_id
join data_merchant on data_merchant.id = data_transaction.id_merchant
join merchant_category on merchant_category.id = data_merchant.id_merchant_category
where card_holder.id = 2 or card_holder.id = 18
"""

fraud_victims = pd.read_sql(query_fraud, engine)
fraud_victims.head()
```

Out[12]:

	id	date	amount
0	18	2018-01-01 23:15:10	2.95
1	18	2018-01-05 07:19:27	1.36
2	2	2018-01-06 02:16:41	1.33
3	2	2018-01-06 05:13:20	10.82
4	18	2018-01-07 01:10:54	175.00

In [ ]:

```
```

In [13]:

```
# id 2 fraud analysis
```

In [14]:

```
victim_two = fraud_victims.loc[fraud_victims['id']== 2]
```

In [15]:

```
victim_two.set_index('id').head()
```

Out[15]:

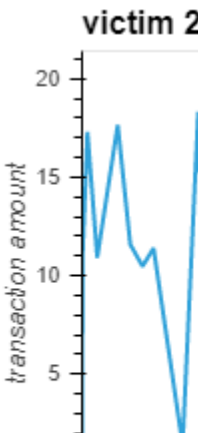
	date	amount
id		
2	2018-01-06 02:16:41	1.33
2	2018-01-06 05:13:20	10.82
2	2018-01-07 15:10:27	17.29
2	2018-01-10 10:07:20	10.91
2	2018-01-16 06:29:35	17.64

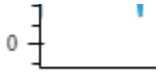
In [16]:

```
plot2 = victim_two.hvplot.line(
    x = 'date',
    y = 'amount',
    xlabel = 'date',
    ylabel = 'transaction amount',
    title = 'victim 2 fraud',
    label = 'card holder 2'
)

plot2
```

Out[16]:





In [ ]:

In [17]:
```
# id 18 fraud analysis
```

In [18]:
```
victim_eighteen = fraud_victims.loc[fraud_victims['id']== 18]
```

In [19]:
```
victim_eighteen.set_index('id').head()
```

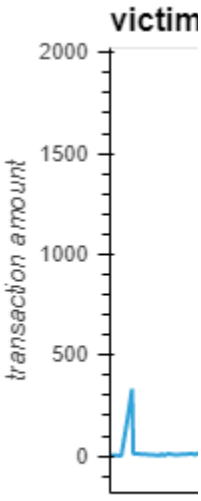
Out[19]:

id	date	amount
18	2018-01-01 23:15:10	2.95
18	2018-01-05 07:19:27	1.36
18	2018-01-07 01:10:54	175.00
18	2018-01-08 11:15:36	333.00
18	2018-01-08 20:10:59	11.55

In [20]:
```
plot18 = victim_eighteen.hvplot.line(
    x = 'date',
    y = 'amount',
    xlabel = 'date',
    ylabel = 'transaction amount',
    title = 'victim 18 fraud',
    label = 'card holder 18'
)

plot18
```

Out[20]:

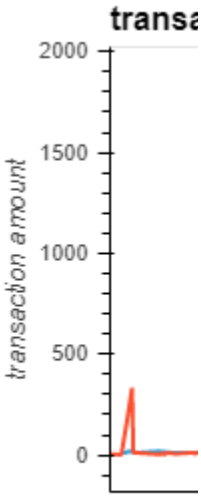


In [ ]:

In [21]:
```
# combined id 2&18 fraud analysis
```

In [22]:
```
combined_plot = plot2 * plot18
combined_plot.opts(title='transaction analysis for card holders 2 and 18')
```

Out[22]:



In [ ]:

In [ ]:

In [ ]:

In [23]:
```
# Data Analysis Question 2
```

In [24]:
```
query_fraud_2 = """
select card_holder.id,
data_transaction.date, data_transaction.amount from data_transaction
join credit_card on credit_card.card = data_transaction.card
join card_holder on card_holder.id = credit_card.cardholder_id
join data_merchant on data_merchant.id = data_transaction.id_merchant
join merchant_category on merchant_category.id = data_merchant.id_merchant_category
where card_holder.id = 25
"""

victim25 = pd.read_sql(query_fraud_2, engine)
victim25.head()
```

Out[24]:

	id	date	amount
0	25	2018-01-02 02:06:21	1.46
1	25	2018-01-05 06:26:45	10.74
2	25	2018-01-07 14:57:23	2.93

	id	date	amount
3	25	2018-01-10 00:25:40	1.39
4	25	2018-01-14 05:02:22	17.84

In [25]:
```
# pull out month from date
victim25['month'] = victim25["date"].apply(lambda x: x.month)
```

In [26]:
```
# pull out day from date
victim25['day'] = victim25["date"].apply(lambda x: x.day)
```

In [27]:
```
# convert month number to month name
victim25['month'] = victim25["date"].apply(lambda x: x.strftime('%B'))
```

In [28]:
```
# drop columns not needed
victim25.drop(columns = ['date', 'id']).head()
```

Out[28]:

	amount	month	day
0	1.46	January	2
1	10.74	January	5
2	2.93	January	7
3	1.39	January	10
4	17.84	January	14

In [29]:
```
# arrange columns
victim25 = victim25[['month', 'day', 'amount']]
victim25.head()
```

Out[29]:

	month	day	amount
0	January	2	1.46
1	January	5	10.74
2	January	7	2.93
3	January	10	1.39
4	January	14	17.84

In [ ]:
```
```

In [34]:
```
# boxplot of victim #25
monthly_analysis = px.box(victim25, x = 'month', y = 'amount', color = 'month')
monthly_analysis
```

In [ ]:

In [40]:
```
# change y-axis to see if data shows better, but doesn't because jsut small transaction
monthly_analysis.update_layout(yaxis_range=[-500,2000])
monthly_analysis.show()
```

In [ ]:

In [ ]:

In [ ]: