Visual Data Analysis of Fraudulent Transactions

Your CFO has also requested detailed trends data on specific card holders. Use the starter notebook to query your database and generate visualizations that supply the requested information as follows, then add your visualizations and observations to your markdown report.

```
In [1]: # Initial imports
        import pandas as pd
        import calendar
        import plotly.express as px
        import hvplot.pandas
        from sqlalchemy import create_engine
```

```
In [2]: # Create a connection to the database
        engine = create_engine("postgresql://postgres:postgres@localhost:5432/fraud_detection")
```

Data Analysis Question 1

The two most important customers of the firm may have been hacked. Verify if there are any fraudulent transactions in their history. For privacy reasons, you only know that their cardholder IDs are 2 and 18.

- Using hvPlot, create a line plot representing the time series of transactions over the course of the year for each cardholder separately.
- Next, to better compare their patterns, create a single line plot that containins both card holders' trend data.
- What difference do you observe between the consumption patterns? Does the difference suggest a fraudulent transaction? Explain your rationale in the markdown report.

```
In [3]: # loading data for card holder 2 and 18 from the database
        # Write the query
        query = """
                YOUR QUERY HERE
                """
        # Create a DataFrame from the query result. HINT: Use pd.read_sql(query, engine)
```
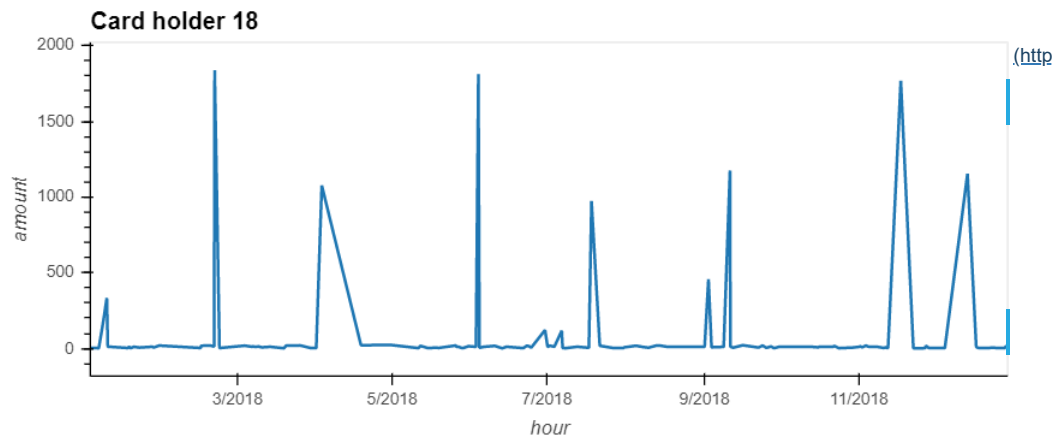
Out[3]:

	cardholder	hour	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 [4]: # Plot for cardholder 2
```

Out[4]:

Card holder 18

(http
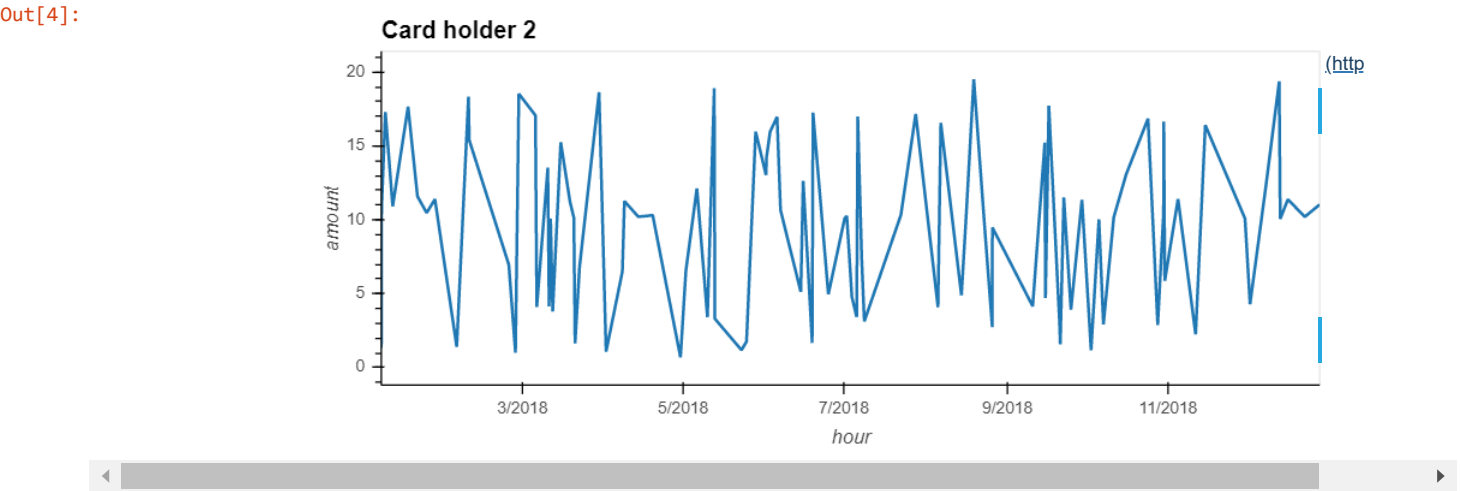
In [6]: # Combined plot for card holders 2 and 18

Out[6]:

Transactions Analysis for Card Holders 2 and 18

(http

Card holder 2
Card holder 18
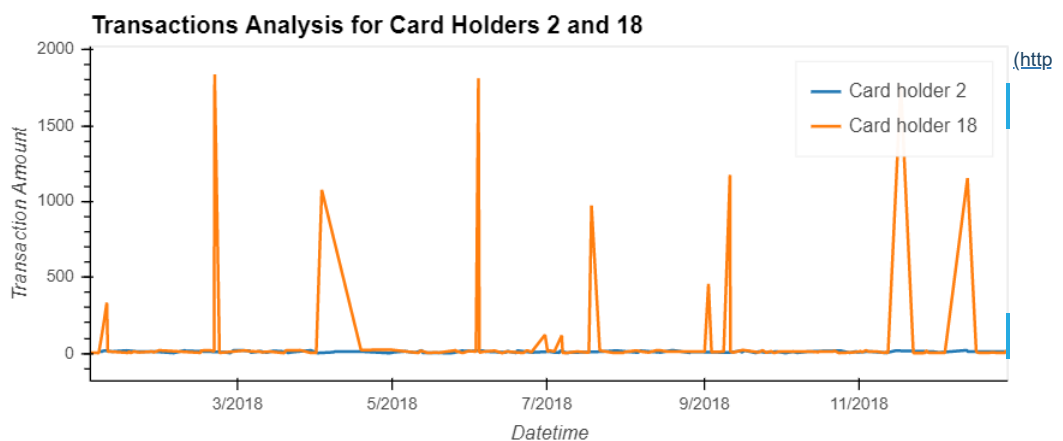
Data Analysis Question 2

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. Again, for privacy reasons, you know only that the cardholder ID in question is 25.

- Using Plotly Express, create a box plot, representing the expenditure data from January 2018 to June 2018 for cardholder ID 25.
- Are there any outliers for cardholder ID 25? How many outliers are there per month?
- Do you notice any anomalies? Describe your observations and conclusions in your markdown report.

```
In [7]: # loading data of daily transactions from jan to jun 2018 for card holder 25
        # Write the query
        query = """
              YOUR QUERY HERE
              """
        # Create a DataFrame from the query result. HINT: Use pd.read_sql(query, engine)
```

Out[7]:

	month	day	amount
0	1.0	2.0	1.46
1	1.0	5.0	10.74
2	1.0	7.0	2.93
3	1.0	10.0	1.39
4	1.0	14.0	17.84

`# loop to change the numeric month to month names`

	month	day	amount
0	January	2.0	1.46
1	January	5.0	10.74
2	January	7.0	2.93
3	January	10.0	1.39
4	January	14.0	17.84

In [9]: `# Creating the six box plots using plotly express`

Monthly Transactions Analysis for Card Holder 25