In [1]:
```
from sqlalchemy import create_engine
import sqlalchemy as db
```

In [4]:
```
engine = db.create_engine('mysql+mysqlconnector://root:testpass123@localhost:
```

In [5]:
```
engine
```

Out[5]:
```
Engine(mysql+mysqlconnector://root:***@localhost:3306/sakila)
```

In [6]:
```
connection = engine.connect()
```

In [7]:
```
connection
```

Out[7]:
```
<sqlalchemy.engine.base.Connection at 0x7facb8327490>
```

In [8]:
```
result = engine.execute('SELECT * FROM actor LIMIT 10')
```

In [9]:
```
result
```

Out[9]:
```
<sqlalchemy.engine.cursor.LegacyCursorResult at 0x7facb8327e80>
```

In [10]:
```
first_result = result.fetchone()
```

In [13]:
```
first_result
```

Out[13]:
```
(1, 'PENELOPE', 'GUINESS', datetime.datetime(2006, 2, 15, 4, 34, 33))
```

In [14]:
```
type(first_result)
```

Out[14]:
```
sqlalchemy.engine.row.LegacyRow
```

In [16]:
```
result.fetchmany(2)
```

Out[16]:
```
[(2, 'NICK', 'WAHLBERG', datetime.datetime(2006, 2, 15, 4, 34, 33)),
 (3, 'ED', 'CHASE', datetime.datetime(2006, 2, 15, 4, 34, 33))]
```

In [18]:
```
other_results = result.fetchall()
```

In [19]:
```
type(other_results)
```

Out[19]:
```
list
```

In [20]:
```
len(other_results)
```

Out[20]:
```
7
```

In [ ]:

In [ ]:

In [ ]:

In [23]:
```
import pandas as pd
```

In [43]:
```
query = 'SELECT * FROM payment'
```

In [44]:
```
posts_df = pd.read_sql_query(query, engine)
```

In [45]:
```
type(posts_df)
```

Out[45]:
```
pandas.core.frame.DataFrame
```

In [46]:
```
posts_df
```

Out[46]:

	payment_id	customer_id	staff_id	rental_id	amount	payment_date	last_update
0	1	1	1	76.0	2.99	2005-05-25 11:30:37	2006-02-15 22:12:30
1	2	1	1	573.0	0.99	2005-05-28 10:35:23	2006-02-15 22:12:30
2	3	1	1	1185.0	5.99	2005-06-15 00:54:12	2006-02-15 22:12:30
3	4	1	2	1422.0	0.99	2005-06-15 18:02:53	2006-02-15 22:12:30
4	5	1	2	1476.0	9.99	2005-06-15 21:08:46	2006-02-15 22:12:30
...	...	...	...	...	...	...	...
16044	16045	599	1	14599.0	4.99	2005-08-21 17:43:42	2006-02-15 22:24:12
16045	16046	599	1	14719.0	1.99	2005-08-21 21:41:57	2006-02-15 22:24:12
16046	16047	599	2	15590.0	8.99	2005-08-23 06:09:44	2006-02-15 22:24:12
16047	16048	599	2	15719.0	2.99	2005-08-23 11:08:46	2006-02-15 22:24:13
16048	16049	599	2	15725.0	2.99	2005-08-23 11:25:00	2006-02-15 22:24:13

16049 rows × 7 columns

In [47]:
```
posts_df.columns
```

Out[47]:
```
Index(['payment_id', 'customer_id', 'staff_id', 'rental_id', 'amount',
       'payment_date', 'last_update'],
      dtype='object')
```

In [48]:
```
posts_df.dtypes
```

Out[48]:
```
payment_id               int64
customer_id              int64
staff_id                 int64
rental_id              float64
amount                 float64
payment_date    datetime64[ns]
last_update     datetime64[ns]
dtype: object
```

In [49]:
```
posts_df.head()
```

Out[49]:

	payment_id	customer_id	staff_id	rental_id	amount	payment_date	last_update
0	1	1	1	76.0	2.99	2005-05-25 11:30:37	2006-02-15 22:12:30

	payment_id	customer_id	staff_id	rental_id	amount	payment_date	last_update
1	2	1	1	573.0	0.99	2005-05-28 10:35:23	2006-02-15 22:12:30
2	3	1	1	1185.0	5.99	2005-06-15 00:54:12	2006-02-15 22:12:30
3	4	1	2	1422.0	0.99	2005-06-15 18:02:53	2006-02-15 22:12:30
4	5	1	2	1476.0	9.99	2005-06-15 21:08:46	2006-02-15 22:12:30

In [50]:
```
posts_df.tail()
```

Out[50]:

	payment_id	customer_id	staff_id	rental_id	amount	payment_date	last_update
16044	16045	599	1	14599.0	4.99	2005-08-21 17:43:42	2006-02-15 22:24:12
16045	16046	599	1	14719.0	1.99	2005-08-21 21:41:57	2006-02-15 22:24:12
16046	16047	599	2	15590.0	8.99	2005-08-23 06:09:44	2006-02-15 22:24:12
16047	16048	599	2	15719.0	2.99	2005-08-23 11:08:46	2006-02-15 22:24:13
16048	16049	599	2	15725.0	2.99	2005-08-23 11:25:00	2006-02-15 22:24:13

In [51]:
```
posts_df.sample(5)
```

Out[51]:

	payment_id	customer_id	staff_id	rental_id	amount	payment_date	last_update
7356	7357	272	2	7658.0	2.99	2005-07-28 02:09:12	2006-02-15 22:15:11
3942	3943	145	2	12785.0	2.99	2005-08-19 00:05:49	2006-02-15 22:13:26
3229	3230	120	1	4001.0	5.99	2005-07-07 00:07:00	2006-02-15 22:13:10
3692	3693	137	1	5804.0	2.99	2005-07-10 15:06:31	2006-02-15 22:13:20
882	883	32	2	2624.0	5.99	2005-06-19 08:22:09	2006-02-15 22:12:36

In [52]:
```
posts_df.describe()
```

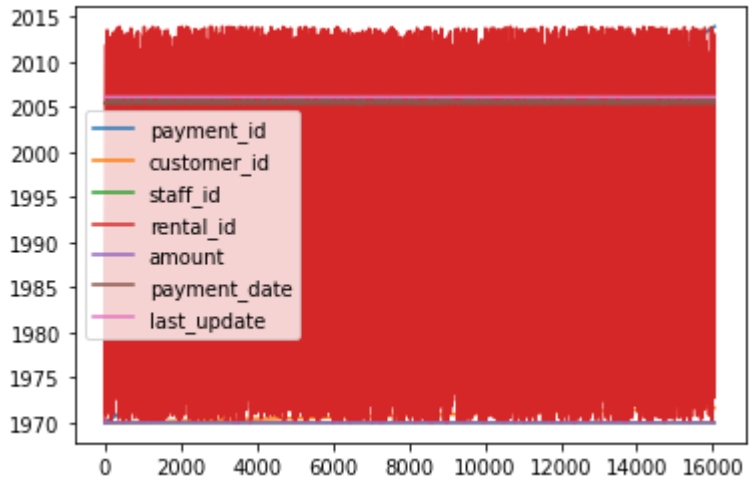
Out[52]:

	payment_id	customer_id	staff_id	rental_id	amount
count	16049.00000	16049.000000	16049.000000	16044.000000	16049.000000
mean	8025.00000	297.162689	1.497975	8025.371478	4.200667
std	4633.09157	172.468100	0.500011	4632.777249	2.362994
min	1.00000	1.000000	1.000000	1.000000	0.000000
25%	4013.00000	148.000000	1.000000	4013.750000	2.990000
50%	8025.00000	296.000000	1.000000	8025.500000	3.990000

	payment_id	customer_id	staff_id	rental_id	amount
75%	12037.00000	446.000000	2.000000	12037.250000	4.990000
max	16049.00000	599.000000	2.000000	16049.000000	11.990000

In [53]:
```
posts_df.plot()
```

Out[53]: <AxesSubplot:>

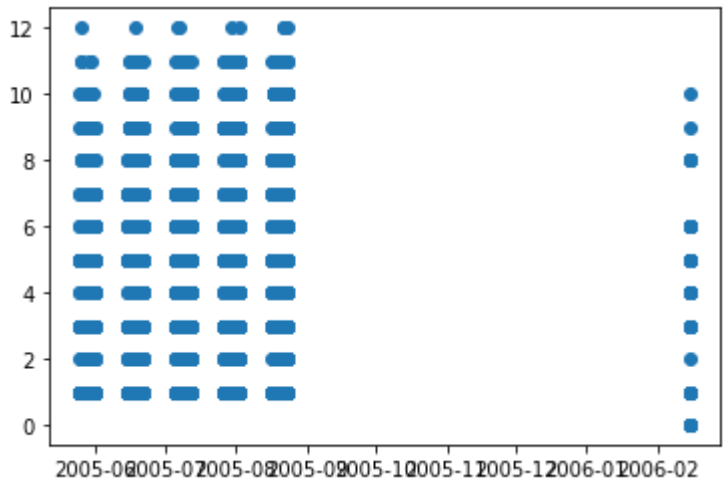


In [54]:
```
import matplotlib.pyplot as plt
```

In [55]:
```
x = posts_df['payment_date']
y = posts_df['amount']
```

In [56]:
```
plt.scatter(x, y)
```

Out[56]: <matplotlib.collections.PathCollection at 0x7facb21fdd60>



In [ ]: