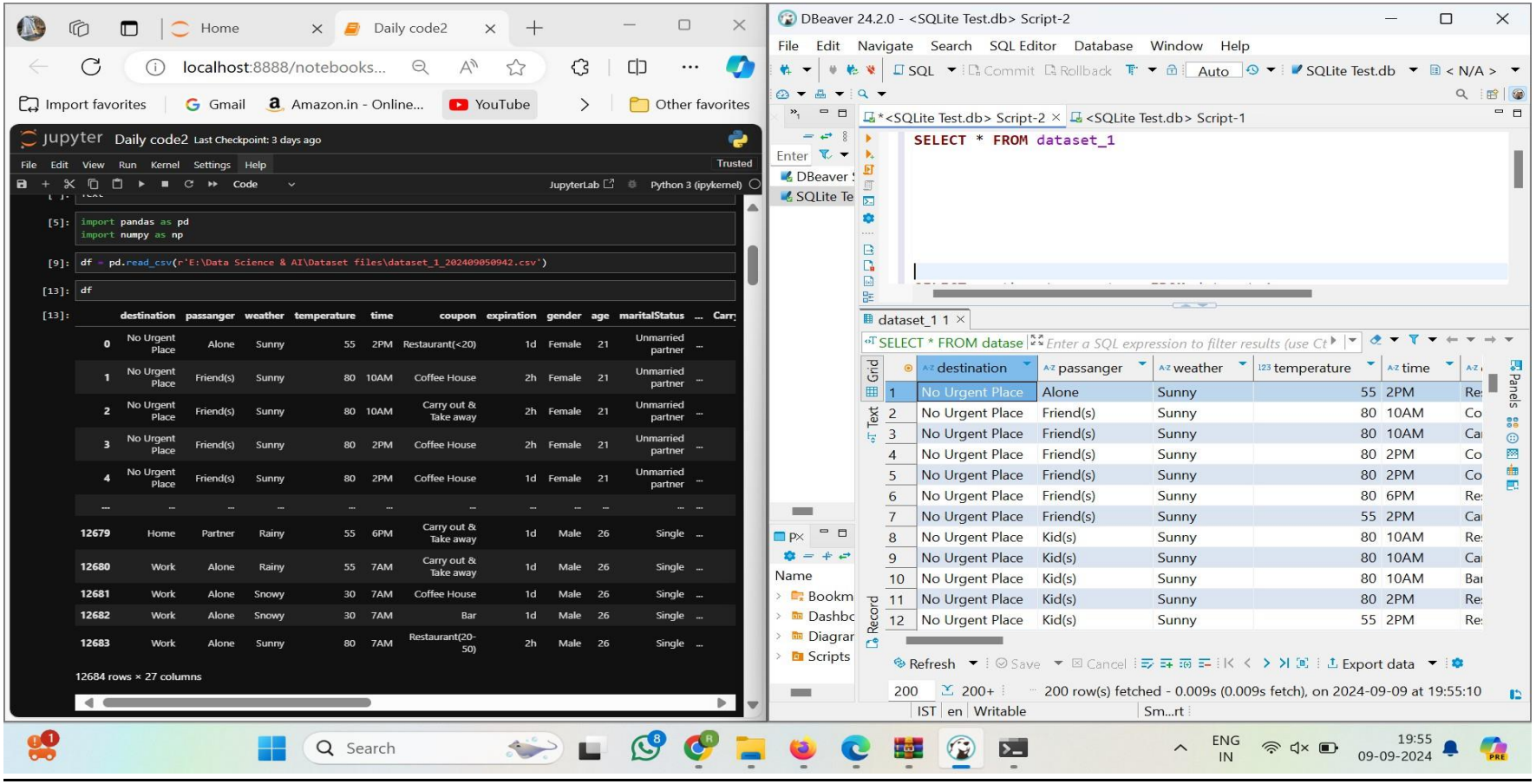
Data Analysis Using SQL vs PYTHON

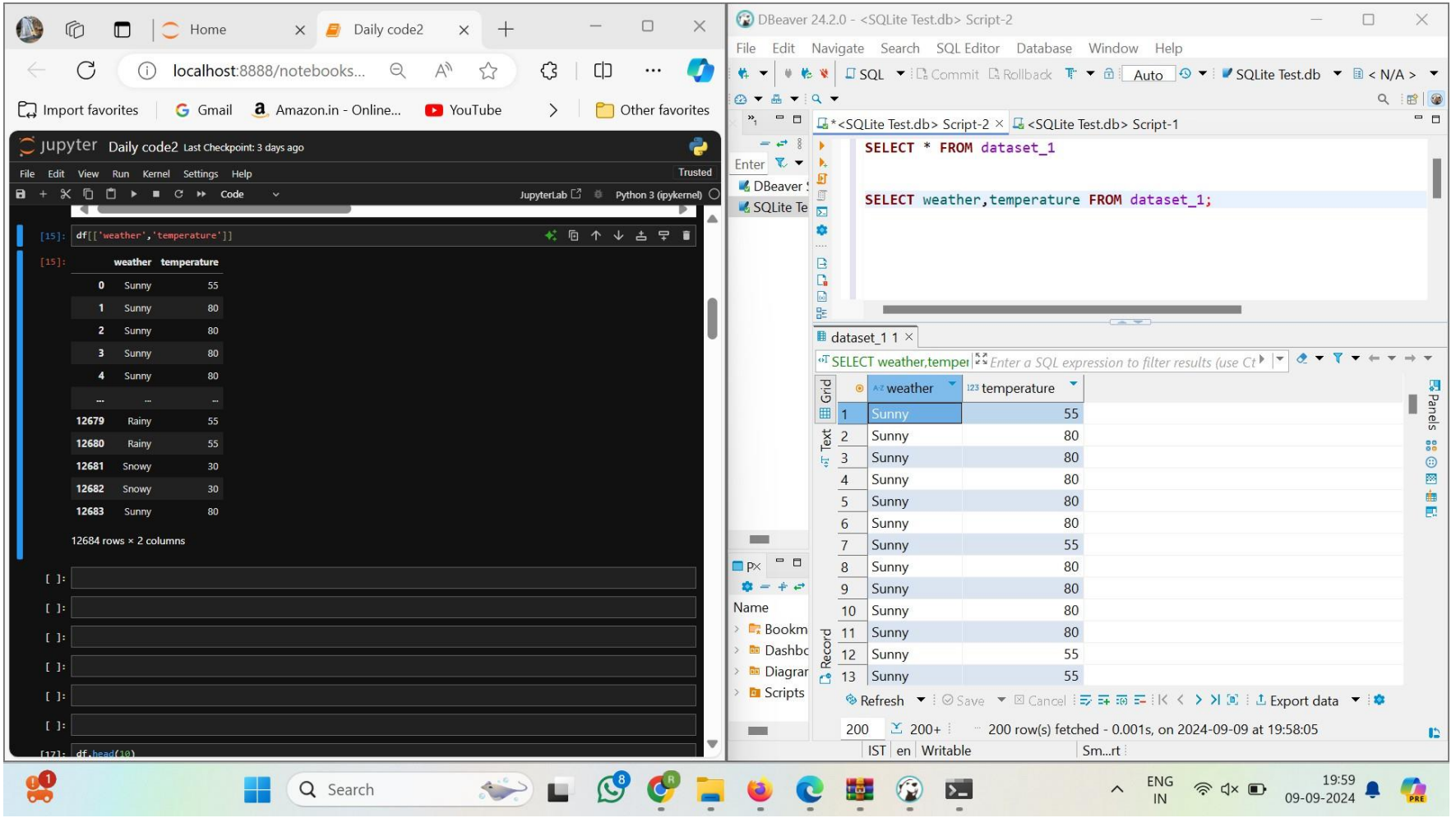
PYTHON	VS	SQL
df = pd.read_csv(r'E:\Data Science & AI\Dataset files\dataset_1_202409050942.csv')		SELECT * FROM dataset_1



PYTHON VS SQL

df[['weather','temperature']]

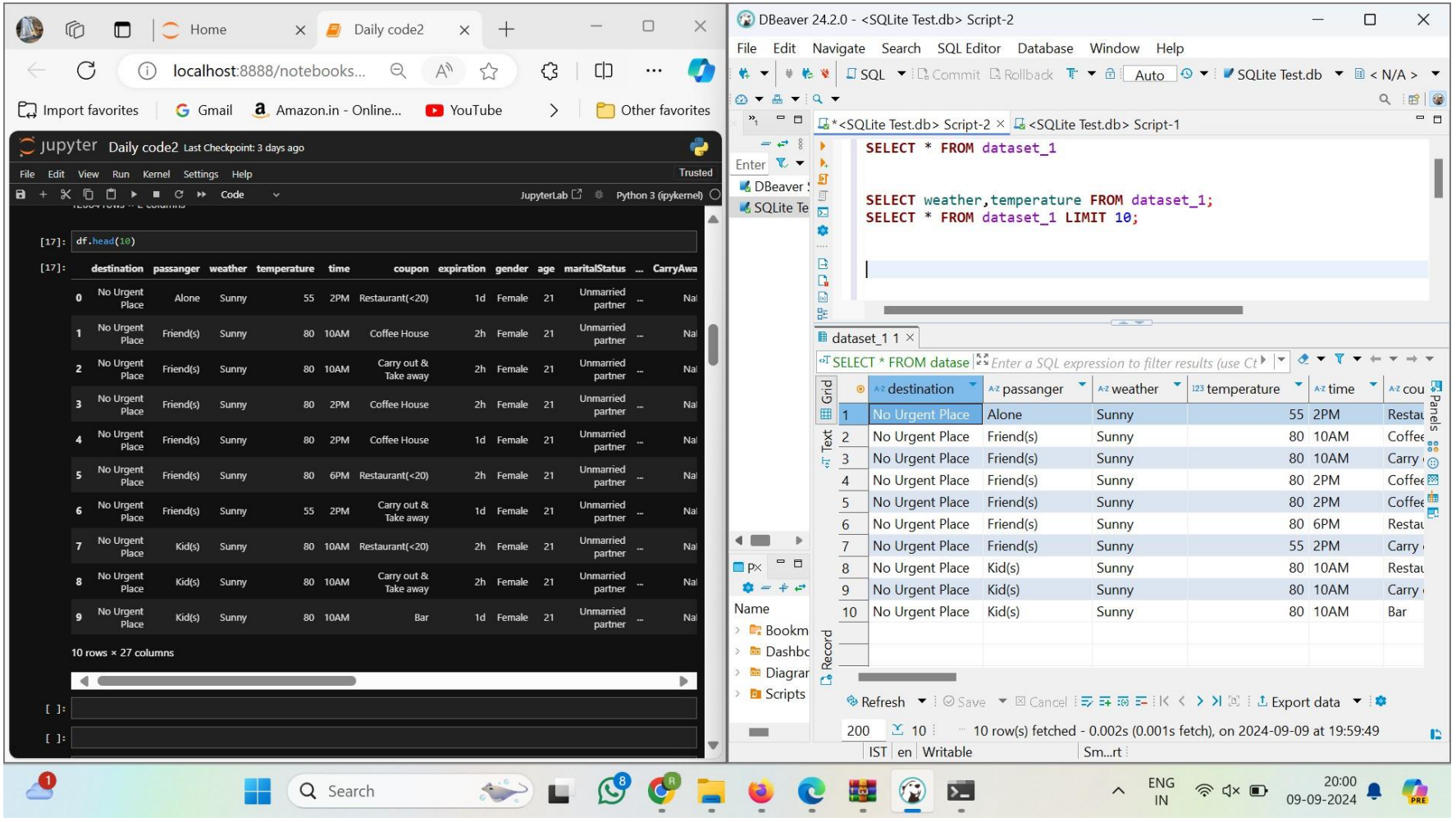
SELECT weather,temperature FROM dataset_1;



PYTHON VS SQL

df.head(10)

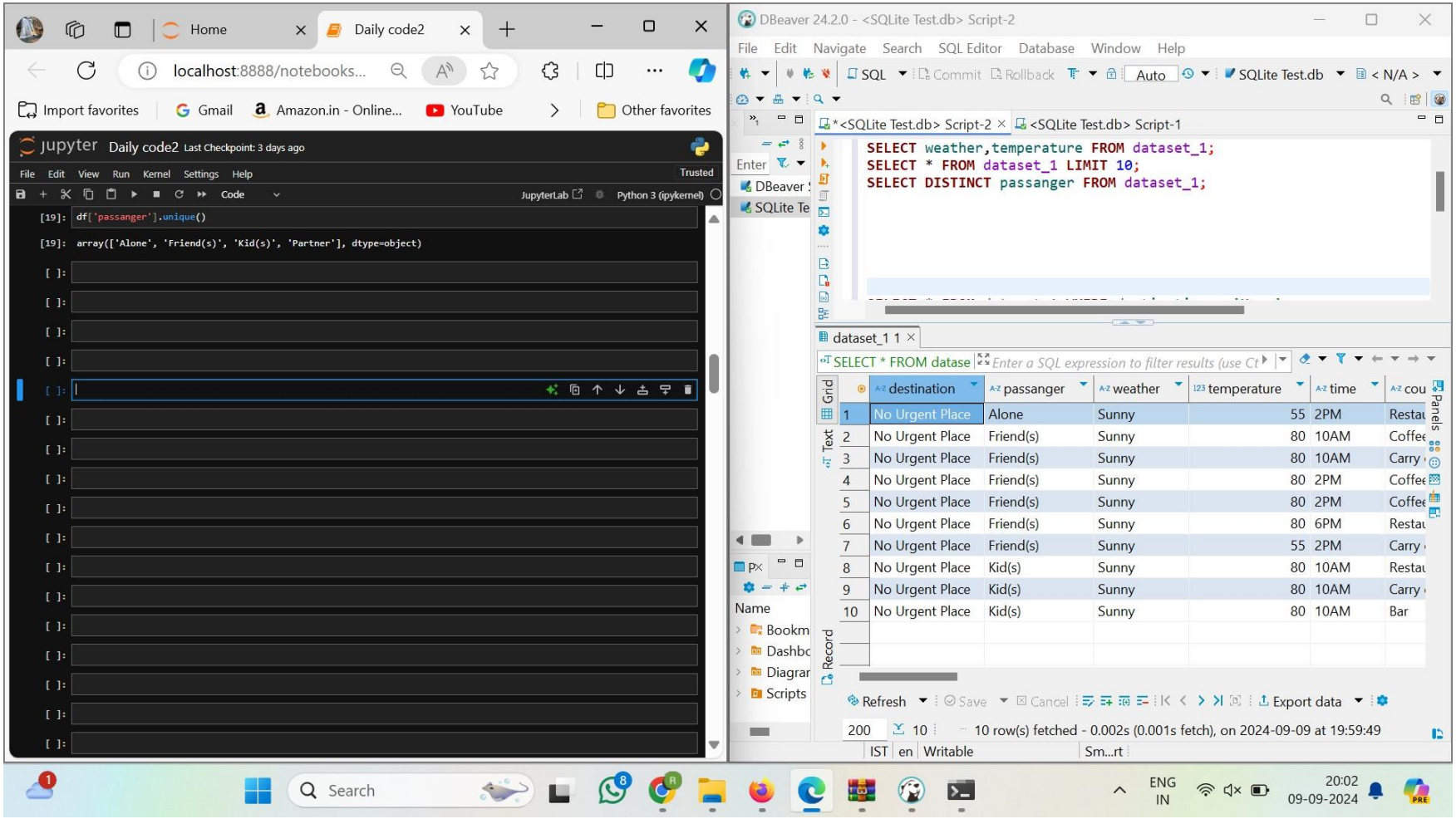
SELECT * FROM dataset_1 LIMIT 10;



PYTHON VS SQL

df['passanger'].unique()

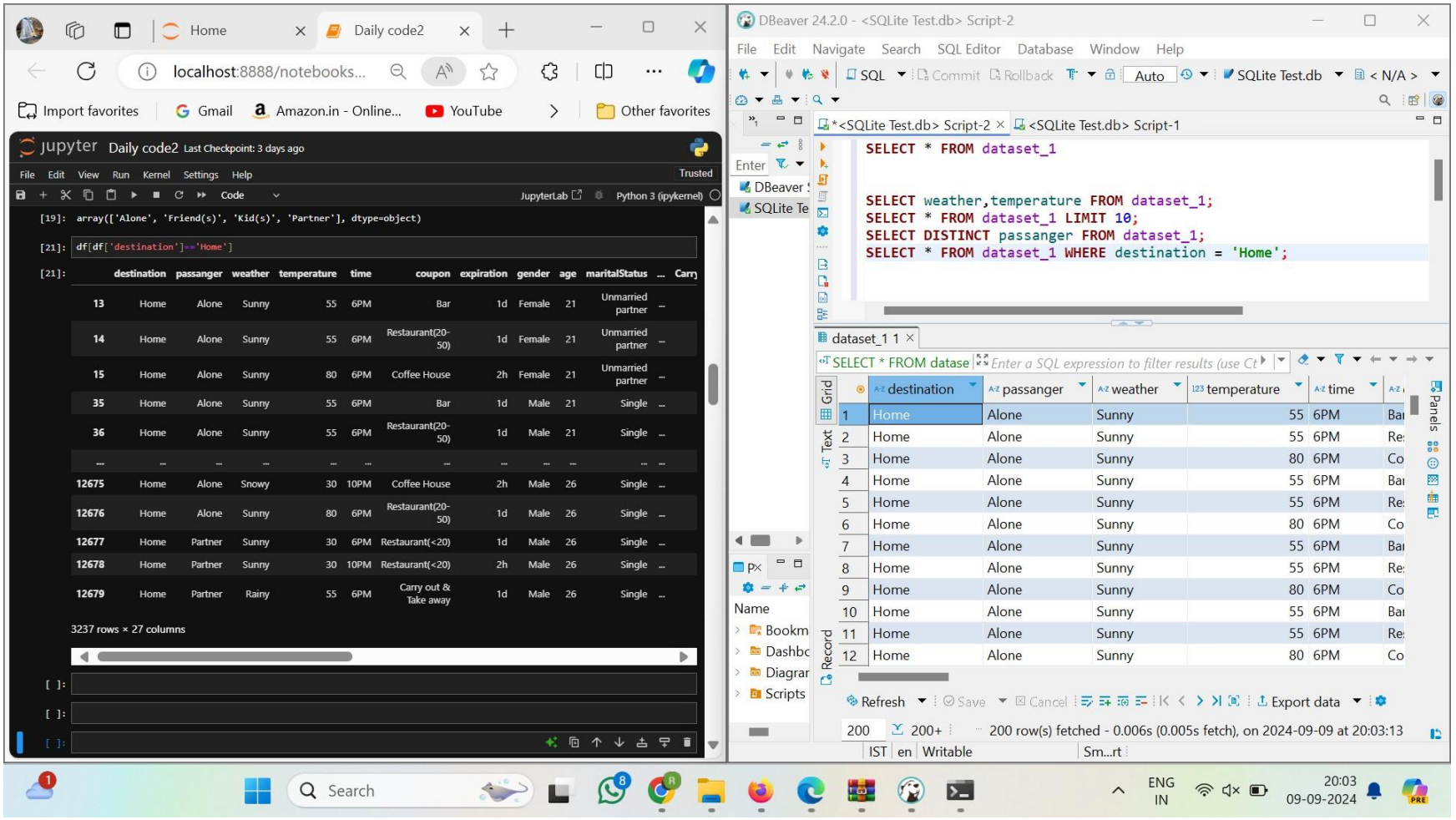
SELECT DISTINCT passanger FROM dataset_1;



PYTHON VS SQL

```
df[df['destination']=='Home']
```

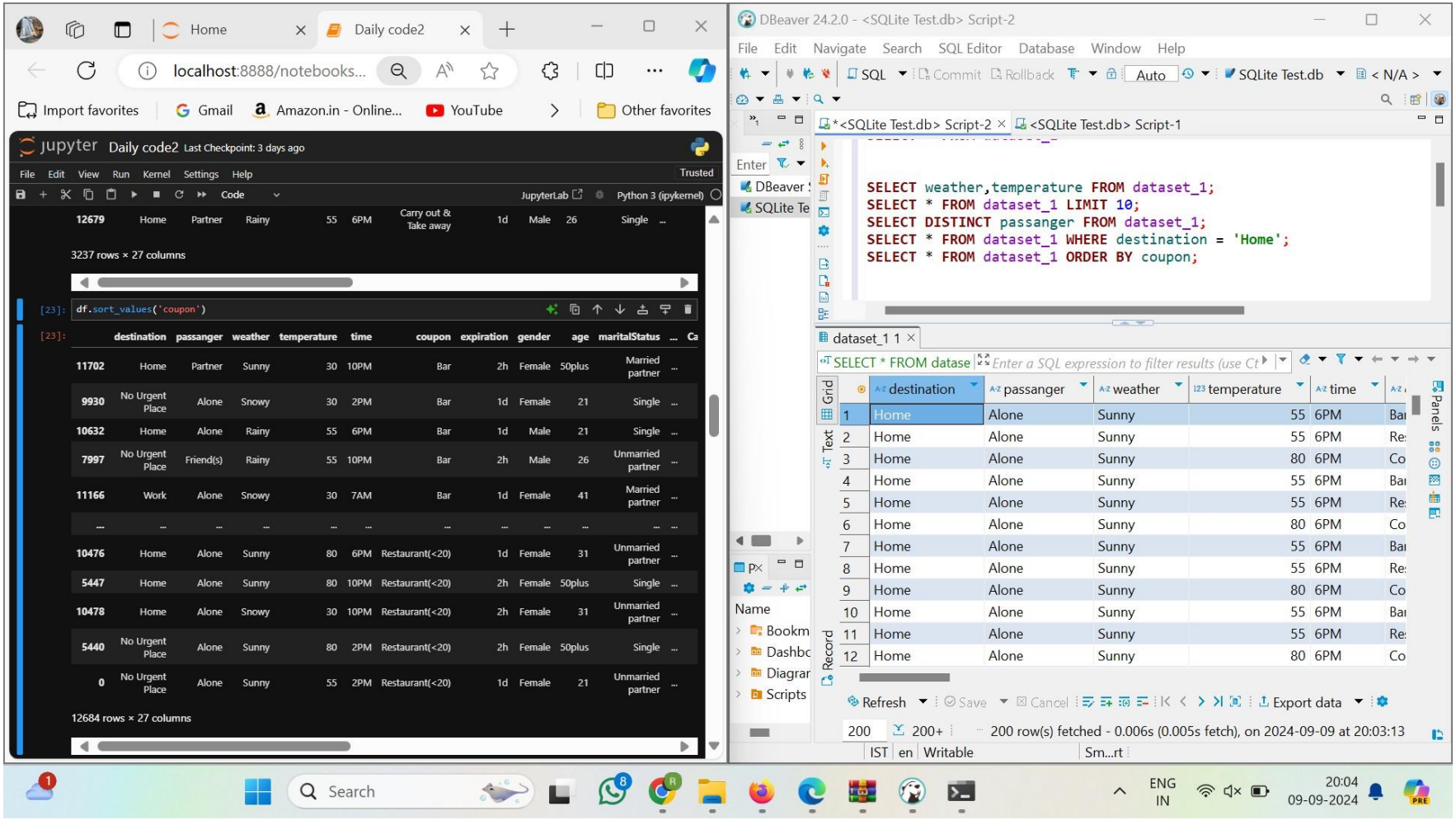
```
SELECT * FROM dataset_1 WHERE destination = 'Home';
```



PYTHON VS SQL

df.sort_values('coupon')

SELECT * FROM dataset_1 ORDER BY coupon;

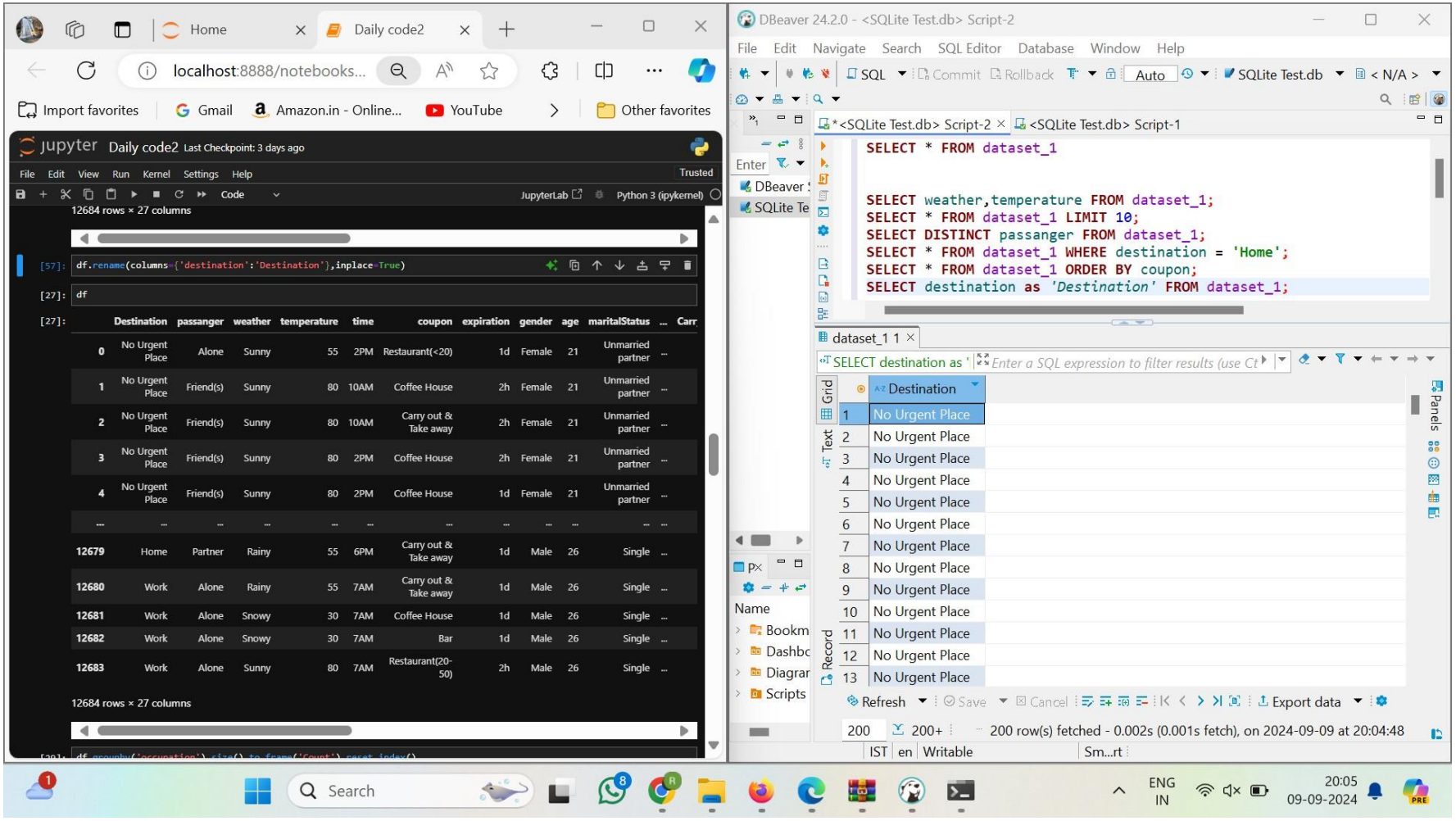


PYTHON VS SQL

df.rename(columns={'destination':'Destination'},inplace=True)

df()

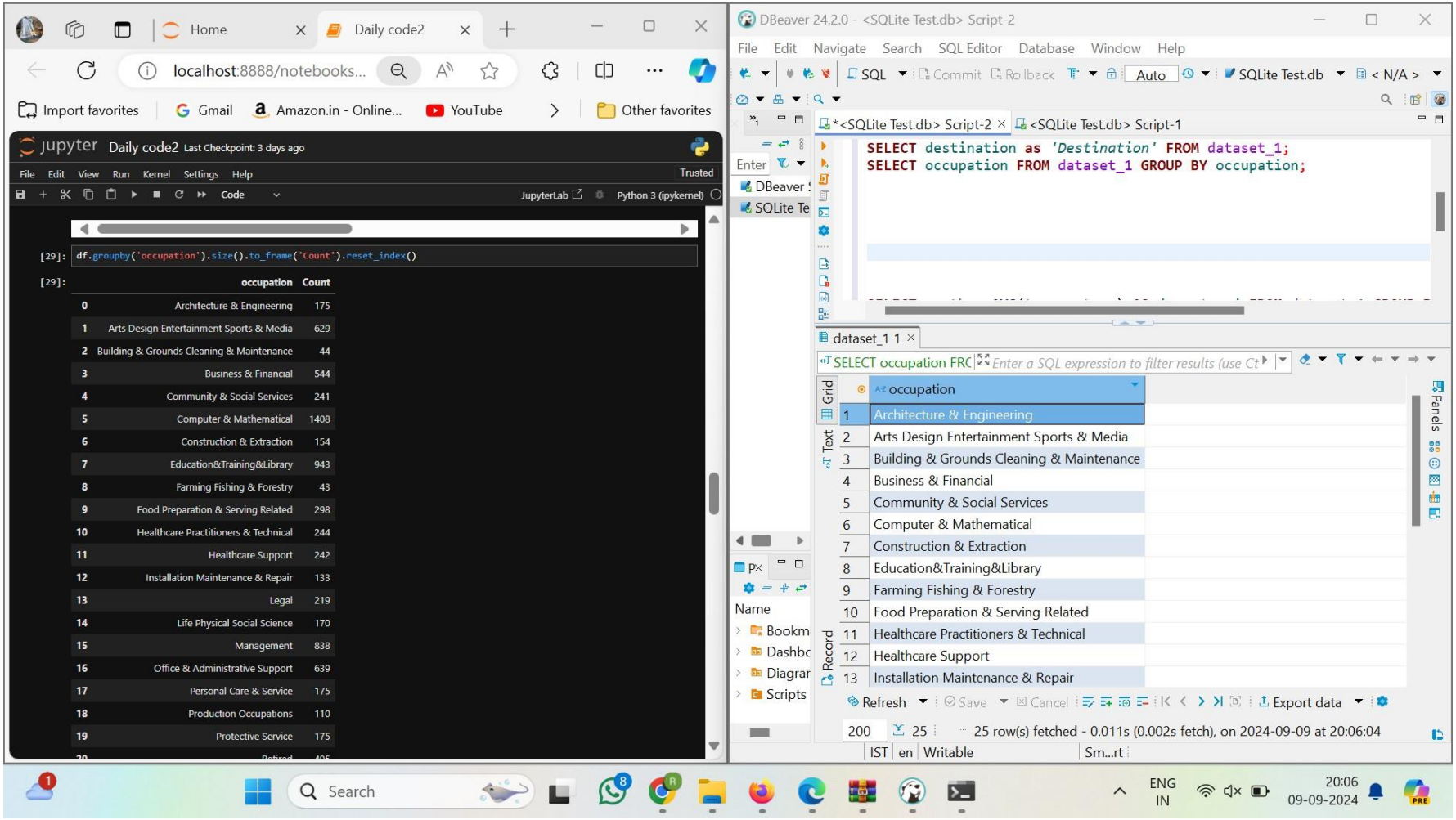
SELECT destination as *'Destination'* FROM dataset_1;



PYTHON VS SQL

df.groupby('occupation').size().to_frame('Count').reset_index()

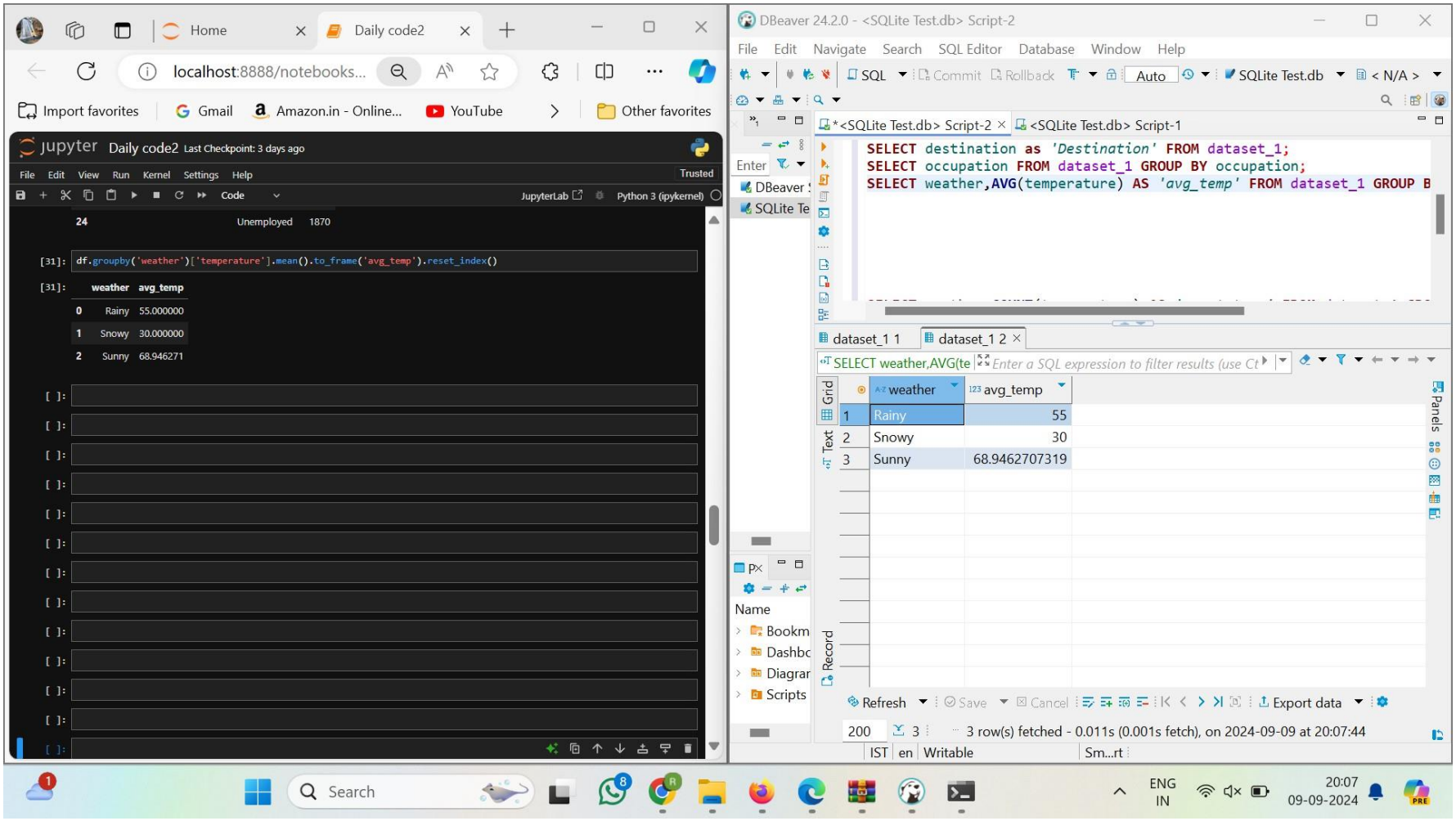
SELECT occupation FROM dataset_1 GROUP BY occupation;



PYTHON VS SQL

```
df.groupby('weather')['temperature'].mean().to_frame
('avg_temp').reset_index()
```

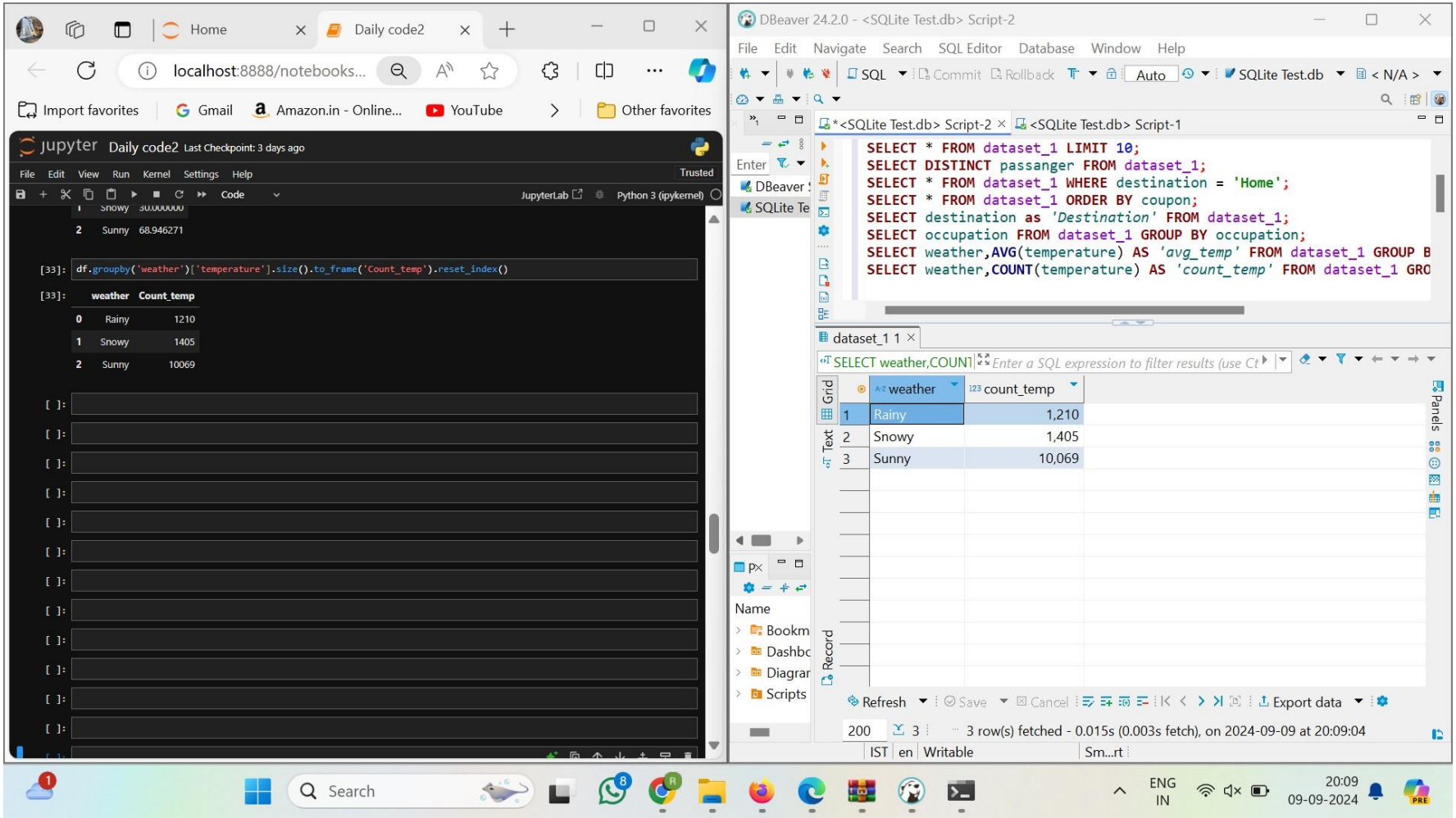
```
SELECT weather,AVG(temperature) AS
'avg_temp' FROM dataset_1 GROUP BY weather;
```



PYTHON VS SQL

```
df.groupby('weather')['temperature'].size().to_frame
('Count_temp').reset_index()
```

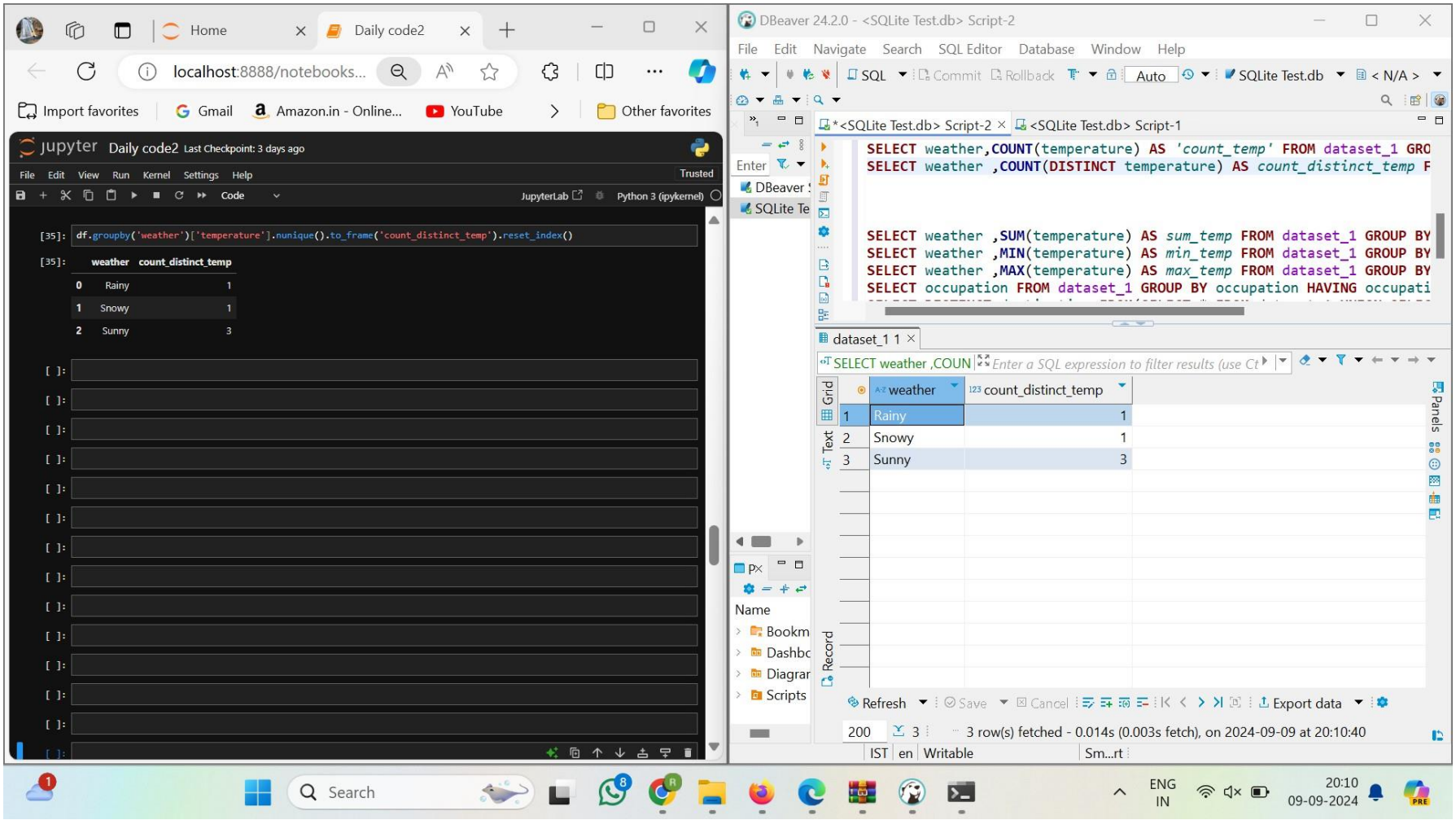
```
SELECT weather,COUNT(temperature) AS
'count_temp' FROM dataset_1 GROUP BY weather;
```



PYTHON VS SQL

```
df.groupby('weather')['temperature'].nunique().to_frame
('count_distinct_temp').reset_index()
```

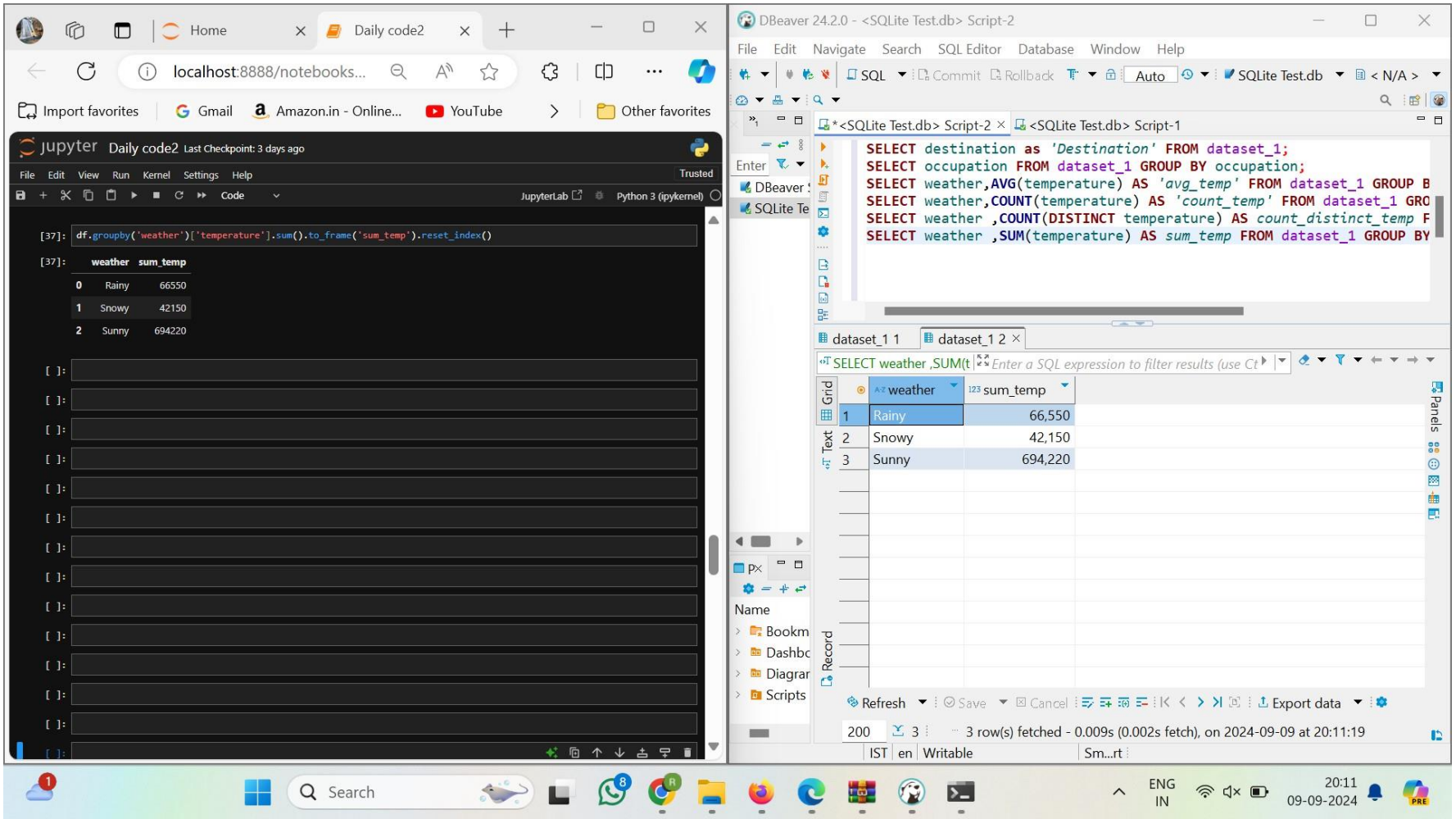
```
SELECT weather ,COUNT(DISTINCT temperature) AS
count_distinct_temp FROM dataset_1 GROUP BY weather;
```



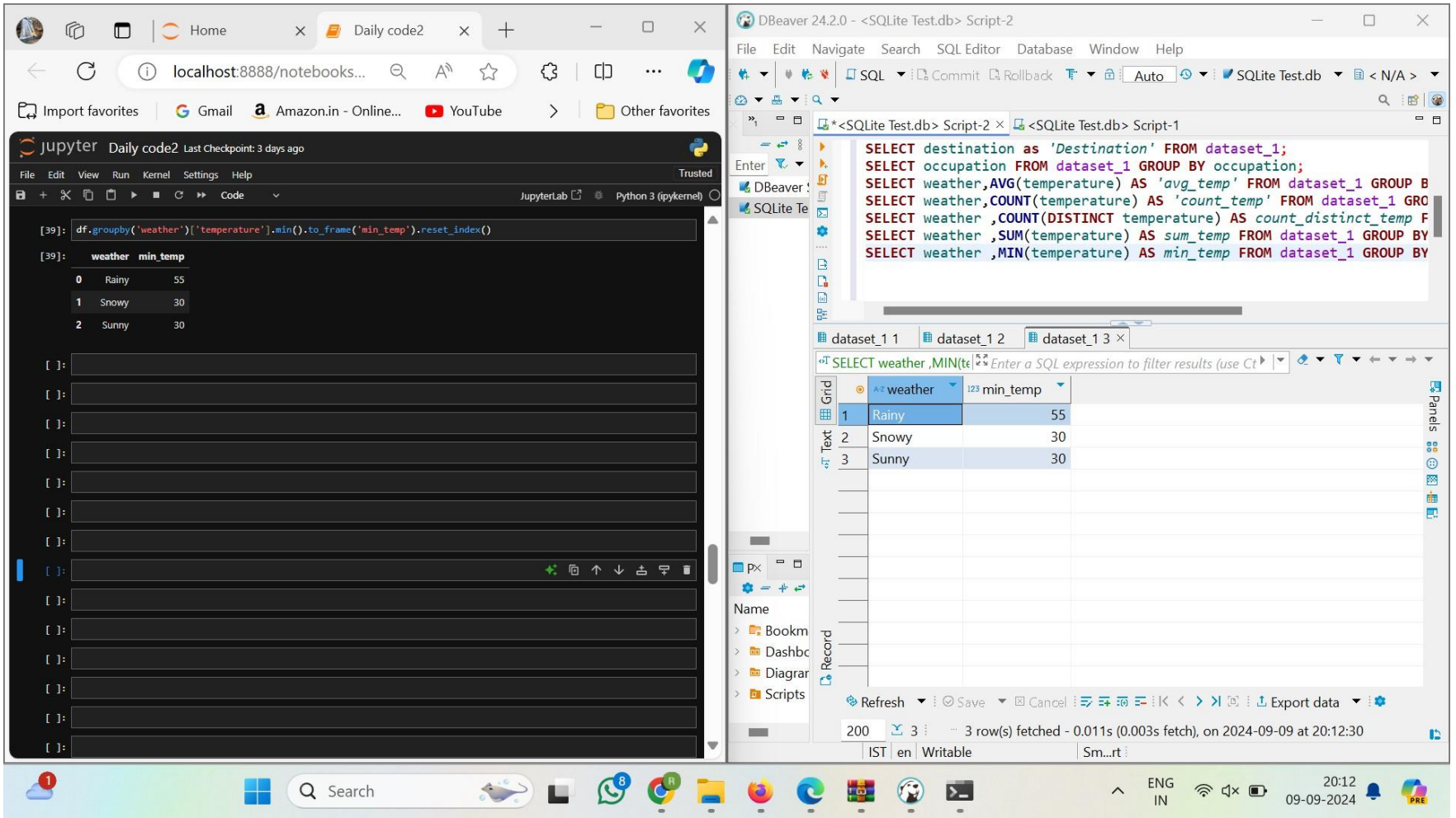
PYTHON VS SQL

```
df.groupby('weather')['temperature'].sum().to_frame
('sum_temp').reset_index()
```

```
SELECT weather ,SUM(temperature) AS sum_temp FROM
dataset_1 GROUP BY weather;
```



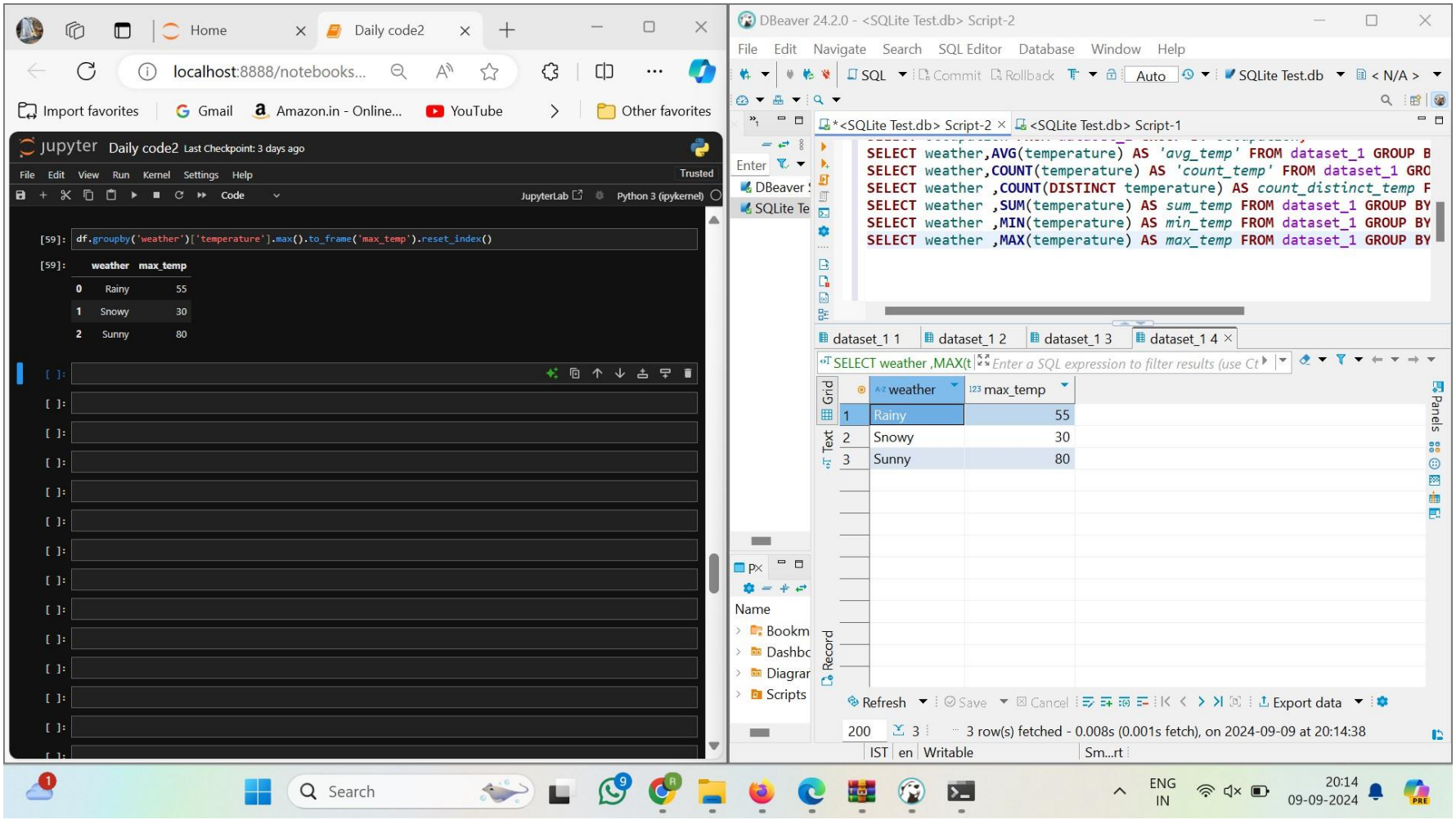
PYTHON	VS	SQL
df.groupby('weather')['temperature'].min().to_frame ('min_temp').reset_index()		SELECT weather ,MIN(temperature) AS *min_temp* FROM dataset_1 GROUP BY weather;



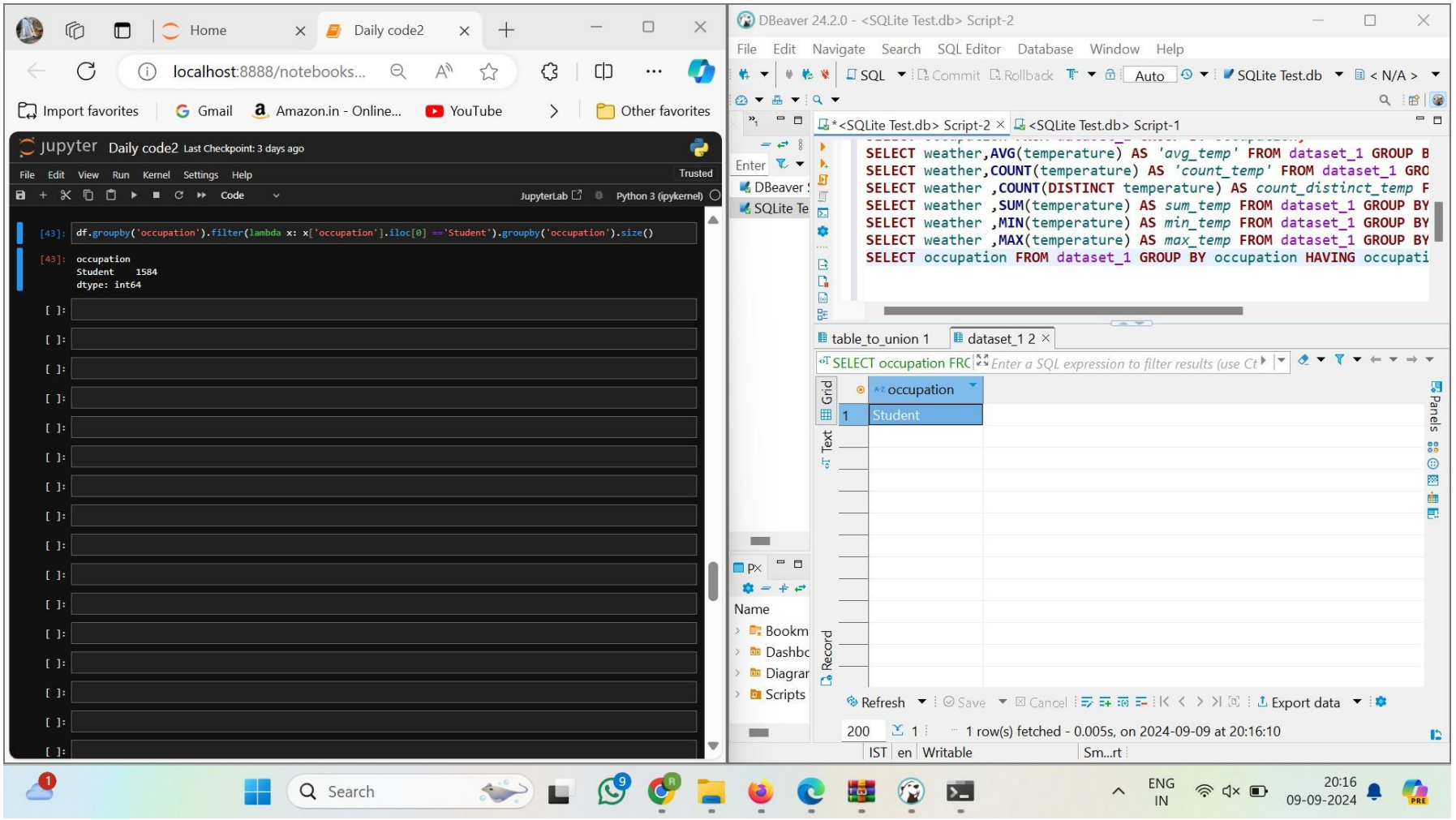
PYTHON	VS	SQL

```
df.groupby('weather')['temperature'].max().to_frame
('max_temp').reset_index()
```

```
SELECT weather ,MAX(temperature) AS
max_temp FROM dataset_1 GROUP BY weather;
```



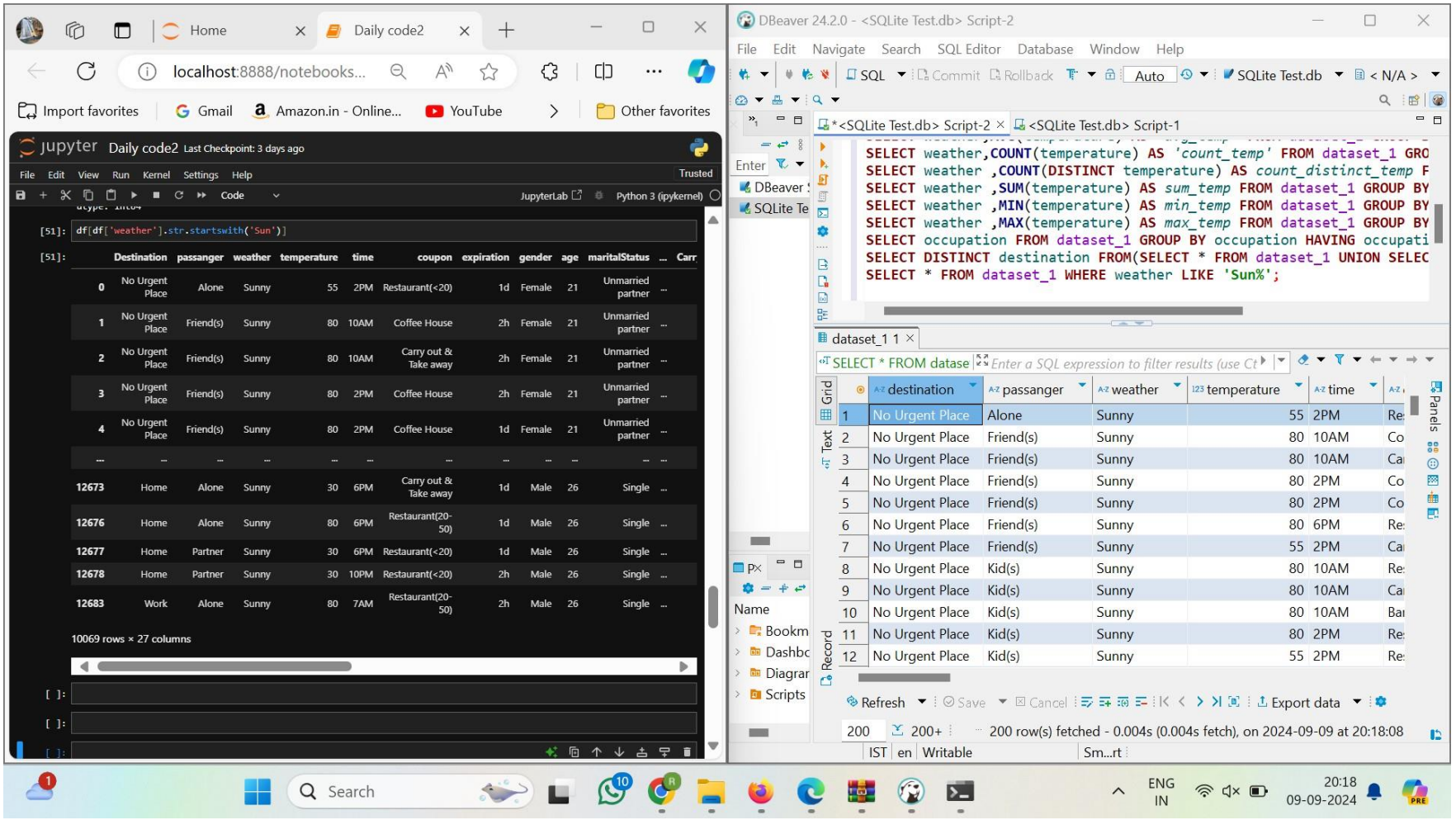
PYTHON	VS	SQL
df.groupby('occupation').filter(lambda x: x['occupation'].iloc[0] == 'Student').groupby('occupation').size()		SELECT occupation FROM dataset_1 GROUP BY occupation HAVING occupation='Student';



PYTHON VS SQL

```
df[df['weather'].str.startswith('Sun')]
```

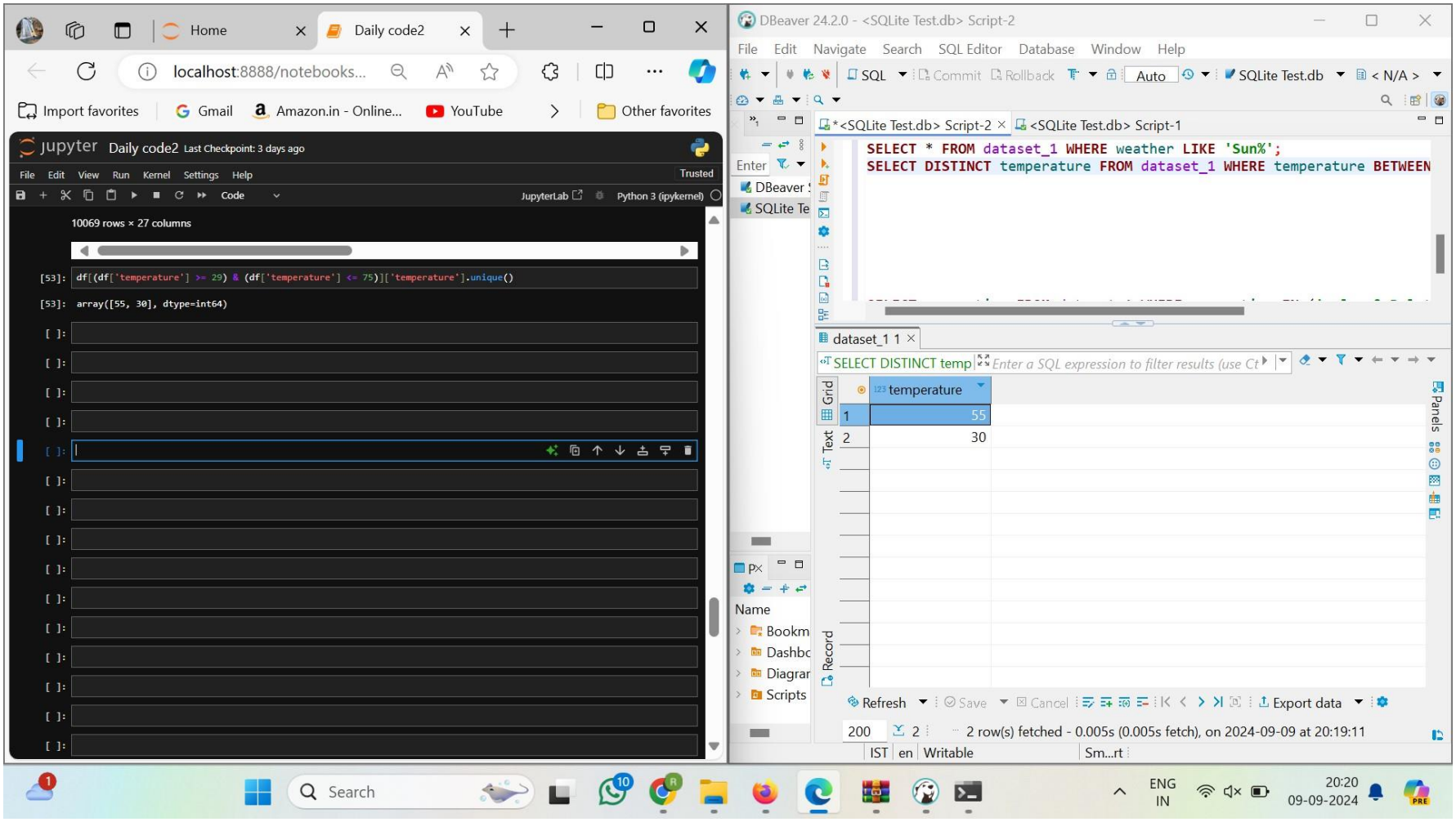
```
SELECT * FROM dataset_1 WHERE weather LIKE 'Sun%';
```



PYTHON VS SQL

```
df[(df['temperature'] >= 29) & (df['temperature'] <= 75)]
['temperature'].unique()
```

```
SELECT DISTINCT temperature FROM
dataset_1 WHERE temperature BETWEEN 29 AND 75;
```



PYTHON VS SQL

```
df[df['occupation'].isin(['Sales & Related', 'Management'])]
[['occupation']]
```

```
SELECT occupation FROM dataset_1
WHERE occupation IN ('sales & Related','Management');
```