Phase 4 – Dataset Merging (Code)

The following Python code was used to merge the cleaned weather and traffic datasets using an inner join on date_time and city.

1. Read Cleaned Datasets

```
import pandas as pd

weather = pd.read_parquet("weather_cleaned.parquet")
traffic = pd.read_parquet("traffic_cleaned.parquet")
```

2. Standardize Merge Keys

```
weather['date_time'] = pd.to_datetime(weather['date_time'])
traffic['date_time'] = pd.to_datetime(traffic['date_time'])

weather['city'] = weather['city'].str.strip().str.lower()
traffic['city'] = traffic['city'].str.strip().str.lower()

weather = weather.dropna(subset=['date_time', 'city'])
traffic = traffic.dropna(subset=['date_time', 'city'])
```

3. Merge Datasets (INNER JOIN)

```
merged = pd.merge(
    weather,
    traffic,
    on=['date_time', 'city'],
    how='inner'
)
```

4. Save Merged Dataset

```
merged.to_parquet("merged_data.parquet")
```