

**SUMMER 2026**

---

**APPENDIX 5***Code for optimised model data preprocessing*

```
# ---- Splitting Training and Testing Data With Shuffling ----

from sklearn.model_selection import train_test_split
import pandas as pd

def split_test_train_val(df, weekcol='Week_ID', testedY=10,
train_test_Y=8, test_ratio=0.25, random_state=1):
    ten_years = df[(df.index.year >= 2015) & (df.index.year <
2025)].copy()

    ten_years[weekcol] = ten_years.index.to_period("W").start_time

    end_date = ten_years.index.max()
    start_date = end_date - pd.DateOffset(years=testedY)

    tenY_df= ten_years[ten_years.index >= start_date].copy()
    valstart=start_date + pd.DateOffset(years=train_test_Y)
    train_test_mask = tenY_df.index < valstart
    valmask = tenY_df.index >= valstart
    train_test_df = tenY_df[train_test_mask].copy()
    val_df = tenY_df[valmask].copy()

    #shuffle weeks
    train_test_weeks= train_test_df[weekcol].unique()
```

**SUMMER 2026**

---

```
trainW, testW= train_test_split(train_test_weeks, test_size=
test_ratio, random_state= random_state)

train_df= train_test_df[train_test_df[weekcol].isin(trainW)]
test_df= train_test_df[train_test_df[weekcol].isin(testW)]
return train_df, test_df, val_df
```