Appendix 3
Exploration and processing of data

[57]:
```
import numpy as np
import pandas as pd
```

1 Cleaning Data

[58]:
```
# Reading orginal dataset into excel file
original_df = pd.read_excel('Original-River-Data.xlsx', usecols='A:I',␣
 ↪skiprows=1)
```

1.0.1 Data Exploration

[59]:
```
original_df.head(10)
```

[59]:

	Unnamed: 0	Crakehill	Skip Bridge	Westwick	Skelton	Arkengarthdale	East Cowton	Malham Tarn	Snaizeholme
0	1993-01-01	10.40	4.393	9.291	26.1	0.0	0	0.0	4.0
1	1993-01-02	9.95	4.239	8.622	24.86	0.0	0	0.8	0.0
2	1993-01-03	9.46	4.124	8.057	23.6	0.0	0	0.8	0.0
3	1993-01-04	9.41	4.363	7.925	23.47	2.4	24.8	0.8	61.6
4	1993-01-05	26.30	11.962	58.704	60.7	11.2	5.6	33.6	111.2
5	1993-01-06	32.10	10.237	34.416	98.01	0.0	0	1.6	0.8
6	1993-01-07	19.30	7.254	22.263	56.99	5.6	4	17.6	36.0
7	1993-01-08	22.00	7.266	29.587	56.66	1.6	0	1.6	2.4
8	1993-01-09	35.50	8.153	60.253	78.1	14.4	0.8	55.2	104.8
9	1993-01-10	51.00	13.276	93.951	125.7	20.8	2.4	76.0	136.8

[60]:
```
original_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1461 entries, 0 to 1460
Data columns (total 9 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Unnamed: 0      1461 non-null   datetime64[ns]
 1   Crakehill       1461 non-null   float64
 2   Skip Bridge     1461 non-null   object
 3   Westwick        1461 non-null   float64
 4   Skelton         1461 non-null   object
 5   Arkengarthdale  1461 non-null   float64
 6   East Cowton     1461 non-null   object
 7   Malham Tarn     1461 non-null   float64
 8   Snaizeholme     1461 non-null   float64
dtypes: datetime64[ns](1), float64(5), object(3)
memory usage: 102.9+ KB
```

[61]:
```
river_data = original_df.copy()
river_data.head()
```

[61]:

	Unnamed: 0	Crakehill	Skip Bridge	Westwick	Skelton	Arkengarthdale	East Cowton	Malham Tarn	Snaizeholme
0	1993-01-01	10.40	4.393	9.291	26.1	0.0	0	0.0	4.0
1	1993-01-02	9.95	4.239	8.622	24.86	0.0	0	0.8	0.0
2	1993-01-03	9.46	4.124	8.057	23.6	0.0	0	0.8	0.0
3	1993-01-04	9.41	4.363	7.925	23.47	2.4	24.8	0.8	61.6
4	1993-01-05	26.30	11.962	58.704	60.7	11.2	5.6	33.6	111.2

[62]:
```
# Renaming Headers
new_columns = {'Unnamed: 0': 'Date'}
new_columns.update({col: f"{col} MDF (Cumecs)" for col in river_data.columns[1:
 ↪5]})
new_columns.update({col: f"{col} DRT (mm)" for col in river_data.columns[5:]})

river_data.rename(
    columns=new_columns,
    inplace=True
)
river_data.head()
```

[62]:
```
        Date  Crakehill MDF (Cumecs) Skip Bridge MDF (Cumecs)  \
0 1993-01-01                   10.40                    4.393
1 1993-01-02                    9.95                    4.239
2 1993-01-03                    9.46                    4.124
3 1993-01-04                    9.41                    4.363
4 1993-01-05                   26.30                   11.962

   Westwick MDF (Cumecs) Skelton MDF (Cumecs)  Arkengarthdale DRT (mm)  \
0                 9.291                 26.1                      0.0
1                 8.622                24.86                      0.0
2                 8.057                 23.6                      0.0
3                 7.925                23.47                      2.4
4                58.704                 60.7                     11.2

  East Cowton DRT (mm)  Malham Tarn DRT (mm)  Snaizeholme DRT (mm)
0                    0                   0.0                   4.0
1                    0                   0.8                   0.0
2                    0                   0.8                   0.0
3                 24.8                   0.8                  61.6
4                  5.6                  33.6                 111.2
```

MDF - Mean Daily Flow

DRT - Daily Rainfall Total

[63]:
```
river_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1461 entries, 0 to 1460
Data columns (total 9 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   Date                     1461 non-null   datetime64[ns]
 1   Crakehill MDF (Cumecs)   1461 non-null   float64
 2   Skip Bridge MDF (Cumecs) 1461 non-null   object
 3   Westwick MDF (Cumecs)    1461 non-null   float64
 4   Skelton MDF (Cumecs)     1461 non-null   object
 5   Arkengarthdale DRT (mm)  1461 non-null   float64
 6   East Cowton DRT (mm)     1461 non-null   object
 7   Malham Tarn DRT (mm)     1461 non-null   float64
 8   Snaizeholme DRT (mm)     1461 non-null   float64
dtypes: datetime64[ns](1), float64(5), object(3)
memory usage: 102.9+ KB
```

[64]:
```
# Converting to non-numeric columns to numeric columns
river_data['Skip Bridge MDF (Cumecs)'] = pd.to_numeric(river_data['Skip Bridge␣
 ↪MDF (Cumecs)'], errors='coerce')
```

```
river_data['Skelton MDF (Cumecs)'] = pd.to_numeric(river_data['Skelton MDF␣
↪(Cumecs)'], errors='coerce')
river_data['East Cowton DRT (mm)'] = pd.to_numeric(river_data['East Cowton DRT␣
↪(mm)'], errors='coerce')
river_data['Date'] = pd.to_datetime(river_data['Date'], errors='coerce')
river_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1461 entries, 0 to 1460
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Date                      1461 non-null   datetime64[ns]
 1   Crakehill MDF (Cumecs)    1461 non-null   float64
 2   Skip Bridge MDF (Cumecs)  1460 non-null   float64
 3   Westwick MDF (Cumecs)     1461 non-null   float64
 4   Skelton MDF (Cumecs)      1460 non-null   float64
 5   Arkengarthdale DRT (mm)   1461 non-null   float64
 6   East Cowton DRT (mm)      1460 non-null   float64
 7   Malham Tarn DRT (mm)      1461 non-null   float64
 8   Snaizeholme DRT (mm)      1461 non-null   float64
dtypes: datetime64[ns](1), float64(8)
memory usage: 102.9 KB
```

[65]:
```
river_data.isna().sum()
```

[65]:
```
Date                        0
Crakehill MDF (Cumecs)      0
Skip Bridge MDF (Cumecs)    1
Westwick MDF (Cumecs)       0
Skelton MDF (Cumecs)        1
Arkengarthdale DRT (mm)     0
East Cowton DRT (mm)        1
Malham Tarn DRT (mm)        0
Snaizeholme DRT (mm)        0
dtype: int64
```

[66]:
```
river_data.describe()
```

[66]:
```
       Crakehill MDF (Cumecs)  Skip Bridge MDF (Cumecs)  \
count             1461.000000               1460.000000
mean                17.519213                  6.929692
std                 52.587125                 28.182097
min               -999.000000               -999.000000
25%                  5.220000                  2.166000
50%                 10.100000                  3.494500
75%                 21.900000                  8.892250
max                220.000000                 80.244000
```

	Westwick MDF (Cumecs)	Skelton MDF (Cumecs)	Arkengarthdale DRT (mm)	\
count	1461.00000	1460.000000	1461.000000	
mean	21.59704	47.093886	8.519233	
std	28.35579	55.712853	133.950452	
min	1.95400	3.694000	-999.000000	
25%	5.67300	12.442500	0.000000	
50%	10.39100	24.260000	0.800000	
75%	26.73600	59.357500	6.400000	
max	374.06100	448.100000	5000.000000	

	East Cowton DRT (mm)	Malham Tarn DRT (mm)	Snaizeholme DRT (mm)
count	1460.000000	1461.000000	1461.000000
mean	11.466301	68.896920	9.725394
std	235.722765	2092.760651	20.984849
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	1.600000	0.800000
75%	5.600000	18.400000	10.400000
max	9000.000000	80000.000000	268.800000

1.0.2 Removing Outliers and Null Values

[67]:
```
# Dropping nulll values
flow_cols = list(river_data.columns[1:5])
rain_cols = list(river_data.columns[5:])

null_values = river_data.isna().any(axis=1)
river_data[null_values] # Rows with at least 1 null value
```

[67]:

	Date	Crakehill MDF (Cumecs)	Skip Bridge MDF (Cumecs)	\
96	1993-04-07	22.9	4.159	
789	1995-03-01	80.6	19.096	
1134	1996-02-09	12.7	NaN	

	Westwick MDF (Cumecs)	Skelton MDF (Cumecs)	Arkengarthdale DRT (mm)	\
96	26.603	NaN	2.4	
789	100.761	174.00	17.6	
1134	7.576	20.58	0.0	

	East Cowton DRT (mm)	Malham Tarn DRT (mm)	Snaizeholme DRT (mm)
96	0.0	0.0	3.2
789	NaN	51.2	21.6
1134	37.6	12.0	14.4

[68]:
```
river_data.dropna(how="any", inplace=True)
river_data[null_values]
```

```
/Users/bhekimaenetja/opt/anaconda3/lib/python3.7/site-
packages/ipykernel_launcher.py:2: UserWarning: Boolean Series key will be
reindexed to match DataFrame index.
```

[68]:
```
Empty DataFrame
Columns: [Date, Crakehill MDF (Cumecs), Skip Bridge MDF (Cumecs), Westwick MDF
(Cumecs), Skelton MDF (Cumecs), Arkengarthdale DRT (mm), East Cowton DRT (mm),
Malham Tarn DRT (mm), Snaizeholme DRT (mm)]
Index: []
```

[69]:
```
# Replacing negative values
river_data[(river_data[flow_cols + rain_cols] < 0).any(1)] # replace all␣
 ↪negative values with zero
```

[69]:

	Date	Crakehill MDF (Cumecs)	Skip Bridge MDF (Cumecs)	\
43	1993-02-13	-999.00	5.476	
73	1993-03-15	-999.00	2.546	
74	1993-03-16	-999.00	2.494	
116	1993-04-27	24.40	11.684	
1203	1996-04-18	7.61	-999.000	

	Westwick MDF (Cumecs)	Skelton MDF (Cumecs)	Arkengarthdale DRT (mm)	\
43	11.051	30.73	0.0	
73	7.179	18.06	11.2	
74	7.232	17.16	0.8	
116	22.181	73.96	-999.0	
1203	13.918	16.84	0.8	

	East Cowton DRT (mm)	Malham Tarn DRT (mm)	Snaizeholme DRT (mm)
43	0.0	0.0	0.0
73	0.0	86.0	19.2
74	0.0	12.8	8.0
116	0.0	0.0	0.0
1203	4.8	47.2	32.0

[70]:
```
river_data[flow_cols+rain_cols] = river_data[flow_cols+rain_cols].
 ↪where((river_data[flow_cols+rain_cols] > -1), 0)
river_data[(river_data[flow_cols + rain_cols] < 0).any(1)]
```

[70]:
```
Empty DataFrame
Columns: [Date, Crakehill MDF (Cumecs), Skip Bridge MDF (Cumecs), Westwick MDF
(Cumecs), Skelton MDF (Cumecs), Arkengarthdale DRT (mm), East Cowton DRT (mm),
Malham Tarn DRT (mm), Snaizeholme DRT (mm)]
Index: []
```

[71]:
```
# Dropping rows with rainfall outliers
rainfall_outliers = river_data[(river_data[rain_cols] > 400).any(1)]
```

```
rainfall_outliers
```

[71]:
```
           Date  Crakehill MDF (Cumecs)  Skip Bridge MDF (Cumecs)  \
771  1995-02-11                    65.0                    31.496
788  1995-02-28                    44.6                    11.563
1104 1996-01-10                    32.4                    10.548

      Westwick MDF (Cumecs)  Skelton MDF (Cumecs)  Arkengarthdale DRT (mm)  \
771                 108.575                136.70                   5000.0
788                  52.105                 93.91                     19.2
1104                 30.086                 84.33                      0.8

      East Cowton DRT (mm)  Malham Tarn DRT (mm)  Snaizeholme DRT (mm)
771                   15.2                 108.4                  80.8
788                 9000.0                  46.4                  47.2
1104                   0.0               80000.0                   0.8
```

[72]:
```
river_data.drop(rainfall_outliers.index, inplace=True)
river_data[(river_data[rain_cols] > 400).any(1)] # drop rows that have rainfall␣
↪outliers
```

[72]:
```
Empty DataFrame
Columns: [Date, Crakehill MDF (Cumecs), Skip Bridge MDF (Cumecs), Westwick MDF
(Cumecs), Skelton MDF (Cumecs), Arkengarthdale DRT (mm), East Cowton DRT (mm),
Malham Tarn DRT (mm), Snaizeholme DRT (mm)]
Index: []
```

[73]:
```
# Dropping rows with river flow outliers
river_flow_outliers = river_data[(river_data[flow_cols] == 0).any(1)]
river_flow_outliers
```

[73]:
```
           Date  Crakehill MDF (Cumecs)  Skip Bridge MDF (Cumecs)  \
43   1993-02-13                    0.00                     5.476
73   1993-03-15                    0.00                     2.546
74   1993-03-16                    0.00                     2.494
1203 1996-04-18                    7.61                     0.000

      Westwick MDF (Cumecs)  Skelton MDF (Cumecs)  Arkengarthdale DRT (mm)  \
43                   11.051                 30.73                      0.0
73                    7.179                 18.06                     11.2
74                    7.232                 17.16                      0.8
1203                 13.918                 16.84                      0.8

      East Cowton DRT (mm)  Malham Tarn DRT (mm)  Snaizeholme DRT (mm)
43                     0.0                   0.0                   0.0
73                     0.0                  86.0                  19.2
74                     0.0                  12.8                   8.0
```

```
1203                 4.8                 47.2                 32.0
```

[74]:
```
river_data.drop(river_flow_outliers.index, inplace=True)
river_data[(river_data[flow_cols] == 0).any(1)]
```

[74]:
```
Empty DataFrame
Columns: [Date, Crakehill MDF (Cumecs), Skip Bridge MDF (Cumecs), Westwick MDF
(Cumecs), Skelton MDF (Cumecs), Arkengarthdale DRT (mm), East Cowton DRT (mm),
Malham Tarn DRT (mm), Snaizeholme DRT (mm)]
Index: []
```

[75]:
```
river_data.describe()
```

[75]:

	Crakehill MDF (Cumecs)	Skip Bridge MDF (Cumecs)	\
count	1451.000000	1451.000000	
mean	19.522233	7.600946	
std	25.249077	10.013759	
min	2.060000	1.002000	
25%	5.240000	2.158500	
50%	10.100000	3.492000	
75%	21.750000	8.833000	
max	220.000000	80.244000	

	Westwick MDF (Cumecs)	Skelton MDF (Cumecs)	Arkengarthdale DRT (mm)	\
count	1451.000000	1451.000000	1451.000000	
mean	21.494272	46.977784	5.784149	
std	28.263025	55.691669	13.262708	
min	1.954000	3.694000	0.000000	
25%	5.632500	12.385000	0.000000	
50%	10.379000	24.250000	0.800000	
75%	26.652000	59.150000	6.400000	
max	374.061000	448.100000	225.200000	

	East Cowton DRT (mm)	Malham Tarn DRT (mm)	Snaizeholme DRT (mm)
count	1451.000000	1451.000000	1451.000000
mean	5.295107	13.986492	9.635837
std	12.296973	25.045850	20.934958
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	1.600000	0.800000
75%	5.600000	18.400000	10.400000
max	165.600000	252.000000	268.800000

1.0.3 Exporting Cleaned Data Set

[76]:
```
export_data = river_data.copy()
export_data["Date"] = export_data["Date"].astype("string")
export_data.to_excel('River-Data-Cleaned.xlsx')
```

2 Lagging and Moving Averages

[77]:
```
clean_df = pd.read_excel('River-Data-Cleaned.xlsx')
clean_df.drop(["Unnamed: 0"], axis=1, inplace=True)
clean_df.head(10)
```

[77]:

	Date	Crakehill MDF (Cumecs)	Skip Bridge MDF (Cumecs)	\
0	1993-01-01	10.40	4.393	
1	1993-01-02	9.95	4.239	
2	1993-01-03	9.46	4.124	
3	1993-01-04	9.41	4.363	
4	1993-01-05	26.30	11.962	
5	1993-01-06	32.10	10.237	
6	1993-01-07	19.30	7.254	
7	1993-01-08	22.00	7.266	
8	1993-01-09	35.50	8.153	
9	1993-01-10	51.00	13.276	

	Westwick MDF (Cumecs)	Skelton MDF (Cumecs)	Arkengarthdale DRT (mm)	\
0	9.291	26.10	0.0	
1	8.622	24.86	0.0	
2	8.057	23.60	0.0	
3	7.925	23.47	2.4	
4	58.704	60.70	11.2	
5	34.416	98.01	0.0	
6	22.263	56.99	5.6	
7	29.587	56.66	1.6	
8	60.253	78.10	14.4	
9	93.951	125.70	20.8	

	East Cowton DRT (mm)	Malham Tarn DRT (mm)	Snaizeholme DRT (mm)
0	0.0	0.0	4.0
1	0.0	0.8	0.0
2	0.0	0.8	0.0
3	24.8	0.8	61.6
4	5.6	33.6	111.2
5	0.0	1.6	0.8
6	4.0	17.6	36.0
7	0.0	1.6	2.4
8	0.8	55.2	104.8
9	2.4	76.0	136.8

[78]:
```
clean_df.describe()
```

[78]:

	Crakehill MDF (Cumecs)	Skip Bridge MDF (Cumecs)	\
count	1451.000000	1451.000000	
mean	19.522233	7.600946	
std	25.249077	10.013759	
min	2.060000	1.002000	
25%	5.240000	2.158500	
50%	10.100000	3.492000	
75%	21.750000	8.833000	
max	220.000000	80.244000	

	Westwick MDF (Cumecs)	Skelton MDF (Cumecs)	Arkengarthdale DRT (mm)	\
count	1451.000000	1451.000000	1451.000000	
mean	21.494272	46.977784	5.784149	
std	28.263025	55.691669	13.262708	
min	1.954000	3.694000	0.000000	
25%	5.632500	12.385000	0.000000	
50%	10.379000	24.250000	0.800000	
75%	26.652000	59.150000	6.400000	
max	374.061000	448.100000	225.200000	

	East Cowton DRT (mm)	Malham Tarn DRT (mm)	Snaizeholme DRT (mm)
count	1451.000000	1451.000000	1451.000000
mean	5.295107	13.986492	9.635837
std	12.296973	25.045850	20.934958
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	1.600000	0.800000
75%	5.600000	18.400000	10.400000
max	165.600000	252.000000	268.800000

[79]:
```
clean_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1451 entries, 0 to 1450
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Date                      1451 non-null   object
 1   Crakehill MDF (Cumecs)    1451 non-null   float64
 2   Skip Bridge MDF (Cumecs)  1451 non-null   float64
 3   Westwick MDF (Cumecs)     1451 non-null   float64
 4   Skelton MDF (Cumecs)      1451 non-null   float64
 5   Arkengarthdale DRT (mm)   1451 non-null   float64
 6   East Cowton DRT (mm)      1451 non-null   float64
 7   Malham Tarn DRT (mm)      1451 non-null   float64
 8   Snaizeholme DRT (mm)      1451 non-null   float64
```

```
dtypes: float64(8), object(1)
memory usage: 102.1+ KB
```

2.0.1 Lagging Data

[80]:
```
lagged_df = pd.DataFrame()
lagged_df["Date"] = clean_df["Date"]
lagged_df[flow_cols[-1]] = clean_df[flow_cols[-1]]
lagged_df.head(10)
```

[80]:

	Date	Skelton MDF (Cumecs)
0	1993-01-01	26.10
1	1993-01-02	24.86
2	1993-01-03	23.60
3	1993-01-04	23.47
4	1993-01-05	60.70
5	1993-01-06	98.01
6	1993-01-07	56.99
7	1993-01-08	56.66
8	1993-01-09	78.10
9	1993-01-10	125.70

[81]:
```
# Lagging rainfall and flow columns by 1 to 3 days
for i in range(3):
    for col in flow_cols:
        col_name = col.replace("(Cumecs)", f"(t-{i+1})")
        lagged_df[col_name] = clean_df[col].shift(i+1)

for i in range(3):
    for col in rain_cols:
        col_name = col.replace("(mm)", f"(t-{i+1})")
        lagged_df[col_name] = clean_df[col].shift(i+1)

lagged_df.head(10)
```

[81]:

	Date	Skelton MDF (Cumecs)	Crakehill MDF (t-1)	\
0	1993-01-01	26.10	NaN	
1	1993-01-02	24.86	10.40	
2	1993-01-03	23.60	9.95	
3	1993-01-04	23.47	9.46	
4	1993-01-05	60.70	9.41	
5	1993-01-06	98.01	26.30	
6	1993-01-07	56.99	32.10	
7	1993-01-08	56.66	19.30	
8	1993-01-09	78.10	22.00	
9	1993-01-10	125.70	35.50	

	Skip Bridge MDF (t-1)	Westwick MDF (t-1)	Skelton MDF (t-1)	\

```
0                    NaN                 NaN                NaN
1                  4.393               9.291              26.10
2                  4.239               8.622              24.86
3                  4.124               8.057              23.60
4                  4.363               7.925              23.47
5                 11.962              58.704              60.70
6                 10.237              34.416              98.01
7                  7.254              22.263              56.99
8                  7.266              29.587              56.66
9                  8.153              60.253              78.10

   Crakehill MDF (t-2)  Skip Bridge MDF (t-2)  Westwick MDF (t-2)  \
0                  NaN                    NaN                 NaN
1                  NaN                    NaN                 NaN
2                10.40                  4.393               9.291
3                 9.95                  4.239               8.622
4                 9.46                  4.124               8.057
5                 9.41                  4.363               7.925
6                26.30                 11.962              58.704
7                32.10                 10.237              34.416
8                19.30                  7.254              22.263
9                22.00                  7.266              29.587

   Skelton MDF (t-2)  …  Malham Tarn DRT (t-1)  Snaizeholme DRT (t-1)  \
0                NaN  …                    NaN                    NaN
1                NaN  …                    0.0                    4.0
2              26.10  …                    0.8                    0.0
3              24.86  …                    0.8                    0.0
4              23.60  …                    0.8                   61.6
5              23.47  …                   33.6                  111.2
6              60.70  …                    1.6                    0.8
7              98.01  …                   17.6                   36.0
8              56.99  …                    1.6                    2.4
9              56.66  …                   55.2                  104.8

   Arkengarthdale DRT (t-2)  East Cowton DRT (t-2)  Malham Tarn DRT (t-2)  \
0                       NaN                    NaN                    NaN
1                       NaN                    NaN                    NaN
2                       0.0                    0.0                    0.0
3                       0.0                    0.0                    0.8
4                       0.0                    0.0                    0.8
5                       2.4                   24.8                    0.8
6                      11.2                    5.6                   33.6
7                       0.0                    0.0                    1.6
8                       5.6                    4.0                   17.6
9                       1.6                    0.0                    1.6
```

```
   Snaizeholme DRT (t-2)  Arkengarthdale DRT (t-3)  East Cowton DRT (t-3)  \
0                    NaN                       NaN                    NaN
1                    NaN                       NaN                    NaN
2                    4.0                       NaN                    NaN
3                    0.0                       0.0                    0.0
4                    0.0                       0.0                    0.0
5                   61.6                       0.0                    0.0
6                  111.2                       2.4                   24.8
7                    0.8                      11.2                    5.6
8                   36.0                       0.0                    0.0
9                    2.4                       5.6                    4.0

   Malham Tarn DRT (t-3)  Snaizeholme DRT (t-3)
0                    NaN                    NaN
1                    NaN                    NaN
2                    NaN                    NaN
3                    0.0                    4.0
4                    0.8                    0.0
5                    0.8                    0.0
6                    0.8                   61.6
7                   33.6                  111.2
8                    1.6                    0.8
9                   17.6                   36.0

[10 rows x 26 columns]
```

[82]:
```
# Dropping rows with null values
lagged_df[lagged_df.isna().any(axis=1)]
```

[82]:
```
         Date  Skelton MDF (Cumecs)  Crakehill MDF (t-1)  \
0  1993-01-01                 26.10                  NaN
1  1993-01-02                 24.86                10.40
2  1993-01-03                 23.60                 9.95

   Skip Bridge MDF (t-1)  Westwick MDF (t-1)  Skelton MDF (t-1)  \
0                    NaN                 NaN                NaN
1                  4.393               9.291              26.10
2                  4.239               8.622              24.86

   Crakehill MDF (t-2)  Skip Bridge MDF (t-2)  Westwick MDF (t-2)  \
0                  NaN                    NaN                 NaN
1                  NaN                    NaN                 NaN
2                 10.4                  4.393               9.291

   Skelton MDF (t-2)  …  Malham Tarn DRT (t-1)  Snaizeholme DRT (t-1)  \
0                NaN  …                    NaN                    NaN
1                NaN  …                    0.0                    4.0
```

```
2               26.1  …                    0.8                    0.0

   Arkengarthdale DRT (t-2)  East Cowton DRT (t-2)  Malham Tarn DRT (t-2)  \
0                       NaN                    NaN                    NaN
1                       NaN                    NaN                    NaN
2                       0.0                    0.0                    0.0

   Snaizeholme DRT (t-2)  Arkengarthdale DRT (t-3)  East Cowton DRT (t-3)  \
0                    NaN                       NaN                    NaN
1                    NaN                       NaN                    NaN
2                    4.0                       NaN                    NaN

   Malham Tarn DRT (t-3)  Snaizeholme DRT (t-3)
0                    NaN                    NaN
1                    NaN                    NaN
2                    NaN                    NaN

[3 rows x 26 columns]
```

[83]:
```
lagged_df.dropna(how="any", inplace=True)
lagged_df[lagged_df.isna().any(axis=1)]
```

[83]:
```
Empty DataFrame
Columns: [Date, Skelton MDF (Cumecs), Crakehill MDF (t-1), Skip Bridge MDF
(t-1), Westwick MDF (t-1), Skelton MDF (t-1), Crakehill MDF (t-2), Skip Bridge
MDF (t-2), Westwick MDF (t-2), Skelton MDF (t-2), Crakehill MDF (t-3), Skip
Bridge MDF (t-3), Westwick MDF (t-3), Skelton MDF (t-3), Arkengarthdale DRT
(t-1), East Cowton DRT (t-1), Malham Tarn DRT (t-1), Snaizeholme DRT (t-1),
Arkengarthdale DRT (t-2), East Cowton DRT (t-2), Malham Tarn DRT (t-2),
Snaizeholme DRT (t-2), Arkengarthdale DRT (t-3), East Cowton DRT (t-3), Malham
Tarn DRT (t-3), Snaizeholme DRT (t-3)]
Index: []

[0 rows x 26 columns]
```

[84]:
```
lagged_df.describe()
```

[84]:
```
       Skelton MDF (Cumecs)  Crakehill MDF (t-1)  Skip Bridge MDF (t-1)  \
count           1448.000000          1448.000000            1448.000000
mean              47.023622            19.537023               7.606707
std               55.740230            25.272741              10.023280
min                3.694000             2.060000               1.002000
25%               12.377500             5.215000               2.154500
50%               24.245000            10.050000               3.480500
75%               59.270000            21.800000               8.855250
max              448.100000           220.000000              80.244000
```

	Westwick MDF (t-1)	Skelton MDF (t-1)	Crakehill MDF (t-2)	\
count	1448.000000	1448.000000	1448.000000	
mean	21.517077	47.012696	19.532983	
std	28.287685	55.743271	25.273807	
min	1.954000	3.694000	2.060000	
25%	5.609250	12.377500	5.215000	
50%	10.385000	24.230000	10.000000	
75%	26.724000	59.270000	21.800000	
max	374.061000	448.100000	220.000000	

	Skip Bridge MDF (t-2)	Westwick MDF (t-2)	Skelton MDF (t-2)	\
count	1448.000000	1448.000000	1448.000000	
mean	7.605369	21.513186	47.006191	
std	10.023601	28.289072	55.745308	
min	1.002000	1.954000	3.694000	
25%	2.154500	5.609250	12.377500	
50%	3.480500	10.354500	24.230000	
75%	8.855250	26.724000	59.270000	
max	80.244000	374.061000	448.100000	

	Crakehill MDF (t-3)	…	Malham Tarn DRT (t-1)	Snaizeholme DRT (t-1)	\
count	1448.000000	…	1448.000000	1448.000000	
mean	19.531809	…	14.010497	9.652486	
std	25.274192	…	25.065902	20.953300	
min	2.060000	…	0.000000	0.000000	
25%	5.215000	…	0.000000	0.000000	
50%	10.000000	…	1.600000	0.800000	
75%	21.800000	…	18.400000	10.400000	
max	220.000000	…	252.000000	268.800000	

	Arkengarthdale DRT (t-2)	East Cowton DRT (t-2)	Malham Tarn DRT (t-2)	\
count	1448.000000	1448.000000	1448.000000	
mean	5.783978	5.295028	14.007182	
std	13.274272	12.308373	25.067332	
min	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	0.000000	
50%	0.800000	0.000000	1.600000	
75%	6.400000	5.300000	18.400000	
max	225.200000	165.600000	252.000000	

	Snaizeholme DRT (t-2)	Arkengarthdale DRT (t-3)	East Cowton DRT (t-3)	\
count	1448.000000	1448.000000	1448.000000	
mean	9.649724	5.781768	5.291713	
std	20.954309	13.274969	12.309152	
min	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	0.000000	
50%	0.800000	0.800000	0.000000	

```
75%                 10.400000                  6.400000                 5.300000
max                268.800000                225.200000               165.600000

       Malham Tarn DRT (t-3)  Snaizeholme DRT (t-3)
count            1448.000000            1448.000000
mean                14.007182               9.651934
std                 25.067332              20.953544
min                  0.000000               0.000000
25%                  0.000000               0.000000
50%                  1.600000               0.800000
75%                 18.400000              10.400000
max                252.000000             268.800000

[8 rows x 25 columns]
```

2.0.2 Moving Averages

[85]:
```
moving_avg_df = pd.DataFrame()
moving_avg_df["Date"] = clean_df["Date"]
moving_avg_df[flow_cols[-1]] = clean_df[flow_cols[-1]]
moving_avg_df.head(10)
```

[85]:

	Date	Skelton MDF (Cumecs)
0	1993-01-01	26.10
1	1993-01-02	24.86
2	1993-01-03	23.60
3	1993-01-04	23.47
4	1993-01-05	60.70
5	1993-01-06	98.01
6	1993-01-07	56.99
7	1993-01-08	56.66
8	1993-01-09	78.10
9	1993-01-10	125.70

[86]:
```
# Creating moving averages of between 3 and 7 days for each numerical column
for i in range(3, 8):
    for col in flow_cols:
        col_name = col.replace("(Cumecs)", f"(MA{i})")
        moving_avg_df[col_name] = clean_df[col].rolling(i).mean()

for i in range(3, 8):
    for col in rain_cols:
        col_name = col.replace("(mm)", f"(MA{i})")
        moving_avg_df[col_name] = clean_df[col].rolling(i).mean()

moving_avg_df.head(10)
```

[86]:

	Date	Skelton MDF (Cumecs)	Crakehill MDF (MA3)	\
0	1993-01-01	26.10	NaN	
1	1993-01-02	24.86	NaN	
2	1993-01-03	23.60	9.936667	
3	1993-01-04	23.47	9.606667	
4	1993-01-05	60.70	15.056667	
5	1993-01-06	98.01	22.603333	
6	1993-01-07	56.99	25.900000	
7	1993-01-08	56.66	24.466667	
8	1993-01-09	78.10	25.600000	
9	1993-01-10	125.70	36.166667	

	Skip Bridge MDF (MA3)	Westwick MDF (MA3)	Skelton MDF (MA3)	\
0	NaN	NaN	NaN	
1	NaN	NaN	NaN	
2	4.252000	8.656667	24.853333	
3	4.242000	8.201333	23.976667	
4	6.816333	24.895333	35.923333	
5	8.854000	33.681667	60.726667	
6	9.817667	38.461000	71.900000	
7	8.252333	28.755333	70.553333	
8	7.557667	37.367667	63.916667	
9	9.565000	61.263667	86.820000	

	Crakehill MDF (MA4)	Skip Bridge MDF (MA4)	Westwick MDF (MA4)	\
0	NaN	NaN	NaN	
1	NaN	NaN	NaN	
2	NaN	NaN	NaN	
3	9.8050	4.27975	8.47375	
4	13.7800	6.17200	20.82700	
5	19.3175	7.67150	27.27550	
6	21.7775	8.45400	30.82700	
7	24.9250	9.17975	36.24250	
8	27.2250	8.22750	36.62975	
9	31.9500	8.98725	51.51350	

	Skelton MDF (MA4)	…	Malham Tarn DRT (MA5)	Snaizeholme DRT (MA5)	\
0	NaN	…	NaN	NaN	
1	NaN	…	NaN	NaN	
2	NaN	…	NaN	NaN	
3	24.5075	…	NaN	NaN	
4	33.1575	…	7.20	35.36	
5	51.4450	…	7.52	34.72	
6	59.7925	…	10.88	41.92	
7	68.0900	…	11.04	42.40	
8	72.4400	…	21.92	51.04	
9	79.3625	…	30.40	56.16	

```
   Arkengarthdale DRT (MA6)  East Cowton DRT (MA6)  Malham Tarn DRT (MA6)  \
0                       NaN                    NaN                    NaN
1                       NaN                    NaN                    NaN
2                       NaN                    NaN                    NaN
3                       NaN                    NaN                    NaN
4                       NaN                    NaN                    NaN
5                  2.266667               5.066667               6.266667
6                  3.200000               5.733333               9.200000
7                  3.466667               5.733333               9.333333
8                  5.866667               5.866667              18.400000
9                  8.933333               2.133333              30.933333

   Snaizeholme DRT (MA6)  Arkengarthdale DRT (MA7)  East Cowton DRT (MA7)  \
0                    NaN                       NaN                    NaN
1                    NaN                       NaN                    NaN
2                    NaN                       NaN                    NaN
3                    NaN                       NaN                    NaN
4                    NaN                       NaN                    NaN
5              29.600000                       NaN                    NaN
6              34.933333                  2.742857               4.914286
7              35.333333                  2.971429               4.914286
8              52.800000                  5.028571               5.028571
9              65.333333                  8.000000               5.371429

   Malham Tarn DRT (MA7)  Snaizeholme DRT (MA7)
0                    NaN                    NaN
1                    NaN                    NaN
2                    NaN                    NaN
3                    NaN                    NaN
4                    NaN                    NaN
5                    NaN                    NaN
6               7.885714              30.514286
7               8.114286              30.285714
8              15.885714              45.257143
9              26.628571              64.800000

[10 rows x 42 columns]
```

[87]:
```
# Dropping rows with null values
moving_avg_df[moving_avg_df.isna().any(axis=1)]
```

[87]:
```
         Date  Skelton MDF (Cumecs)  Crakehill MDF (MA3)  \
0  1993-01-01                 26.10                  NaN
1  1993-01-02                 24.86                  NaN
2  1993-01-03                 23.60             9.936667
3  1993-01-04                 23.47             9.606667
```

```
4  1993-01-05                60.70            15.056667
5  1993-01-06                98.01            22.603333

   Skip Bridge MDF (MA3)  Westwick MDF (MA3)  Skelton MDF (MA3)  \
0                    NaN                 NaN                NaN
1                    NaN                 NaN                NaN
2               4.252000            8.656667          24.853333
3               4.242000            8.201333          23.976667
4               6.816333           24.895333          35.923333
5               8.854000           33.681667          60.726667

   Crakehill MDF (MA4)  Skip Bridge MDF (MA4)  Westwick MDF (MA4)  \
0                  NaN                    NaN                 NaN
1                  NaN                    NaN                 NaN
2                  NaN                    NaN                 NaN
3               9.8050                4.27975             8.47375
4              13.7800                6.17200            20.82700
5              19.3175                7.67150            27.27550

   Skelton MDF (MA4)  …  Malham Tarn DRT (MA5)  Snaizeholme DRT (MA5)  \
0                NaN  …                    NaN                    NaN
1                NaN  …                    NaN                    NaN
2                NaN  …                    NaN                    NaN
3           24.5075  …                    NaN                    NaN
4           33.1575  …                   7.20                  35.36
5           51.4450  …                   7.52                  34.72

   Arkengarthdale DRT (MA6)  East Cowton DRT (MA6)  Malham Tarn DRT (MA6)  \
0                       NaN                    NaN                    NaN
1                       NaN                    NaN                    NaN
2                       NaN                    NaN                    NaN
3                       NaN                    NaN                    NaN
4                       NaN                    NaN                    NaN
5                  2.266667               5.066667               6.266667

   Snaizeholme DRT (MA6)  Arkengarthdale DRT (MA7)  East Cowton DRT (MA7)  \
0                    NaN                       NaN                    NaN
1                    NaN                       NaN                    NaN
2                    NaN                       NaN                    NaN
3                    NaN                       NaN                    NaN
4                    NaN                       NaN                    NaN
5                   29.6                       NaN                    NaN

   Malham Tarn DRT (MA7)  Snaizeholme DRT (MA7)
0                    NaN                    NaN
1                    NaN                    NaN
2                    NaN                    NaN
```

```
3                    NaN                    NaN
4                    NaN                    NaN
5                    NaN                    NaN

[6 rows x 42 columns]
```

[88]:
```
moving_avg_df.dropna(how="any", inplace=True)
moving_avg_df[moving_avg_df.isna().any(axis=1)]
```

[88]:
```
Empty DataFrame
Columns: [Date, Skelton MDF (Cumecs), Crakehill MDF (MA3), Skip Bridge MDF
(MA3), Westwick MDF (MA3), Skelton MDF (MA3), Crakehill MDF (MA4), Skip Bridge
MDF (MA4), Westwick MDF (MA4), Skelton MDF (MA4), Crakehill MDF (MA5), Skip
Bridge MDF (MA5), Westwick MDF (MA5), Skelton MDF (MA5), Crakehill MDF (MA6),
Skip Bridge MDF (MA6), Westwick MDF (MA6), Skelton MDF (MA6), Crakehill MDF
(MA7), Skip Bridge MDF (MA7), Westwick MDF (MA7), Skelton MDF (MA7),
Arkengarthdale DRT (MA3), East Cowton DRT (MA3), Malham Tarn DRT (MA3),
Snaizeholme DRT (MA3), Arkengarthdale DRT (MA4), East Cowton DRT (MA4), Malham
Tarn DRT (MA4), Snaizeholme DRT (MA4), Arkengarthdale DRT (MA5), East Cowton DRT
(MA5), Malham Tarn DRT (MA5), Snaizeholme DRT (MA5), Arkengarthdale DRT (MA6),
East Cowton DRT (MA6), Malham Tarn DRT (MA6), Snaizeholme DRT (MA6),
Arkengarthdale DRT (MA7), East Cowton DRT (MA7), Malham Tarn DRT (MA7),
Snaizeholme DRT (MA7)]
Index: []

[0 rows x 42 columns]
```

[89]:
```
moving_avg_df.describe()
```

[89]:
```
       Skelton MDF (Cumecs)  Crakehill MDF (MA3)  Skip Bridge MDF (MA3)  \
count           1445.000000          1445.000000            1445.000000
mean              46.995172            19.545218               7.608666
std               55.777355            23.160818               9.371964
min                3.694000             2.083333               1.012667
25%               12.370000             5.383333               2.176333
50%               24.240000            10.276667               3.568667
75%               59.050000            23.966667               9.565000
max              448.100000           176.666667              67.550333

       Westwick MDF (MA3)  Skelton MDF (MA3)  Crakehill MDF (MA4)  \
count         1445.000000        1445.000000          1445.000000
mean            21.515503          47.028298            19.546846
std             25.103715          52.756141            22.399448
min              1.990333           3.733000             2.095000
25%              5.825667          12.820000             5.585000
50%             10.934333          24.773333            10.435000
75%             29.075333          62.006667            25.050000
```

```
max            251.526667         361.066667           158.250000

       Skip Bridge MDF (MA4)  Westwick MDF (MA4)  Skelton MDF (MA4)  \
count            1445.000000         1445.000000        1445.000000
mean                7.609420           21.521166          47.033714
std                 9.116826           24.211841          51.530965
min                 1.014750            2.003750           3.834250
25%                 2.179500            5.850000          13.050000
50%                 3.593250           11.279000          25.365000
75%                 9.582250           29.579000          63.012500
max                65.921750          216.440250         337.300000

       Crakehill MDF (MA5)  …  Malham Tarn DRT (MA5)  Snaizeholme DRT (MA5)  \
count          1445.000000  …            1445.000000            1445.000000
mean             19.547596  …              14.027183               9.613619
std              21.768351  …              15.169586              12.909387
min               2.110000  …               0.000000               0.000000
25%               5.696000  …               1.760000               0.800000
50%              10.806000  …               9.760000               4.960000
75%              25.380000  …              21.120000              13.280000
max             157.960000  …              90.240000             105.040000

       Arkengarthdale DRT (MA6)  East Cowton DRT (MA6)  Malham Tarn DRT (MA6)  \
count                1445.000000            1445.000000           1.445000e+03
mean                    5.792203               5.292826           1.402713e+01
std                     7.301313               6.078797           1.446201e+01
min                     0.000000               0.000000           8.659740e-15
25%                     0.666667               0.933333           2.133333e+00
50%                     3.466667               3.466667           1.033333e+01
75%                     8.400000               7.266667           2.140000e+01
max                    73.400000              40.400000           9.733333e+01

       Snaizeholme DRT (MA6)  Arkengarthdale DRT (MA7)  East Cowton DRT (MA7)  \
count            1445.000000              1.445000e+03            1445.000000
mean                9.621961              5.791162e+00               5.291626
std                12.375209              6.983505e+00               5.732045
min                 0.000000              5.075305e-16               0.000000
25%                 1.200000              8.000000e-01               1.257143
50%                 5.333333              3.771429e+00               3.657143
75%                13.333333              8.342857e+00               7.142857
max                87.800000              6.965714e+01              35.200000

       Malham Tarn DRT (MA7)  Snaizeholme DRT (MA7)
count           1.445000e+03            1445.000000
mean            1.402709e+01               9.628314
std             1.387434e+01              11.911054
min             5.773160e-15               0.000000
```

```
25%             2.171429e+00              1.371429
50%             1.051429e+01              5.828571
75%             2.102857e+01             13.371429
max             1.040571e+02             75.257143

[8 rows x 41 columns]
```

[90]:
```
# Creating lagged moving averages
lagged_ma_df = pd.DataFrame()
lagged_ma_df["Date"] = moving_avg_df["Date"]
lagged_ma_df[flow_cols[-1]] = moving_avg_df[flow_cols[-1]]
lagged_ma_df.head(10)
```

[90]:

	Date	Skelton MDF (Cumecs)
6	1993-01-07	56.99
7	1993-01-08	56.66
8	1993-01-09	78.10
9	1993-01-10	125.70
10	1993-01-11	195.90
11	1993-01-12	125.40
12	1993-01-13	161.50
13	1993-01-14	204.00
14	1993-01-15	200.60
15	1993-01-16	234.40

[91]:
```
## Lagging moving averages by 1 day
## lagging them by more than 1 day results in much weaker correlations
mdf_cols = list(moving_avg_df.columns[2:22])
drt_cols = list(moving_avg_df.columns[22:])

for col in mdf_cols:
    col_name = col + f" (t-1)"
    lagged_ma_df[col_name] = moving_avg_df[col].shift(1)

for col in drt_cols:
    col_name = col + f" (t-1)"
    lagged_ma_df[col_name] = moving_avg_df[col].shift(1)

lagged_ma_df.head(10)
```

[91]:

	Date	Skelton MDF (Cumecs)	Crakehill MDF (MA3) (t-1)	\
6	1993-01-07	56.99	NaN	
7	1993-01-08	56.66	25.900000	
8	1993-01-09	78.10	24.466667	
9	1993-01-10	125.70	25.600000	
10	1993-01-11	195.90	36.166667	
11	1993-01-12	125.40	50.666667	

```
12  1993-01-13               161.50                  49.500000
13  1993-01-14               204.00                  53.933333
14  1993-01-15               200.60                  55.133333
15  1993-01-16               234.40                  76.000000

    Skip Bridge MDF (MA3) (t-1)  Westwick MDF (MA3) (t-1)  \
6                           NaN                       NaN
7                      9.817667                 38.461000
8                      8.252333                 28.755333
9                      7.557667                 37.367667
10                     9.565000                 61.263667
11                    15.663333                 74.569000
12                    19.850667                 67.989333
13                    27.605333                 72.256667
14                    28.570667                 69.197333
15                    32.919333                103.079000

    Skelton MDF (MA3) (t-1)  Crakehill MDF (MA4) (t-1)  \
6                       NaN                       NaN
7                 71.900000                   21.7775
8                 70.553333                   24.9250
9                 63.916667                   27.2250
10                86.820000                   31.9500
11               133.233333                   43.5000
12               149.000000                   46.0000
13               160.933333                   53.2000
14               163.633333                   57.7250
15               188.700000                   65.0000

    Skip Bridge MDF (MA4) (t-1)  Westwick MDF (MA4) (t-1)  \
6                           NaN                       NaN
7                       8.45400                  30.82700
8                       9.17975                  36.24250
9                       8.22750                  36.62975
10                      8.98725                  51.51350
11                     13.56400                  63.32350
12                     16.92625                  66.05525
13                     24.02300                  77.68025
14                     27.81825                  69.27375
15                     29.86825                  87.43775

    Skelton MDF (MA4) (t-1)  …  Malham Tarn DRT (MA5) (t-1)  \
6                       NaN  …                         NaN
7                   59.7925  …                       10.88
8                   68.0900  …                       11.04
9                   72.4400  …                       21.92
10                  79.3625  …                       30.40
```

```
11                114.0900  …                        32.48
12                131.2750  …                        29.12
13                152.1250  …                        45.36
14                171.7000  …                        35.92
15                172.8750  …                        35.12

    Snaizeholme DRT (MA5) (t-1)  Arkengarthdale DRT (MA6) (t-1)  \
6                           NaN                            NaN
7                         41.92                       3.200000
8                         42.40                       3.466667
9                         51.04                       5.866667
10                        56.16                       8.933333
11                        61.60                       8.800000
12                        59.20                      10.000000
13                        82.32                      15.600000
14                        63.92                      16.266667
15                        50.96                      19.866667

    East Cowton DRT (MA6) (t-1)  Malham Tarn DRT (MA6) (t-1)  \
6                           NaN                           NaN
7                      5.733333                      9.200000
8                      5.733333                      9.333333
9                      5.866667                     18.400000
10                     2.133333                     30.933333
11                     3.866667                     27.333333
12                     4.533333                     27.200000
13                     5.333333                     38.066667
14                     5.600000                     39.133333
15                     7.466667                     41.933333

    Snaizeholme DRT (MA6) (t-1)  Arkengarthdale DRT (MA7) (t-1)  \
6                           NaN                            NaN
7                     34.933333                       2.742857
8                     35.333333                       2.971429
9                     52.800000                       5.028571
10                    65.333333                       8.000000
11                    51.466667                       9.142857
12                    55.333333                       8.571429
13                    69.000000                      14.171429
14                    70.733333                      14.171429
15                    65.266667                      19.085714

    East Cowton DRT (MA7) (t-1)  Malham Tarn DRT (MA7) (t-1)  \
6                           NaN                           NaN
7                      4.914286                      7.885714
8                      4.914286                      8.114286
9                      5.028571                     15.885714
```

```
10                     5.371429                     26.628571
11                     4.114286                     28.228571
12                     3.885714                     23.542857
13                     5.142857                     35.142857
14                     4.800000                     33.771429
15                     6.514286                     43.828571

    Snaizeholme DRT (MA7) (t-1)
6                           NaN
7                     30.514286
8                     30.285714
9                     45.257143
10                    64.800000
11                    60.000000
12                    47.542857
13                    64.285714
14                    60.971429
15                    70.914286

[10 rows x 42 columns]
```

[92]:
```
# Dropping rows with null values
lagged_ma_df[lagged_ma_df.isna().any(axis=1)]
```

[92]:
```
         Date  Skelton MDF (Cumecs)  Crakehill MDF (MA3) (t-1)  \
6  1993-01-07                 56.99                        NaN

   Skip Bridge MDF (MA3) (t-1)  Westwick MDF (MA3) (t-1)  \
6                          NaN                       NaN

   Skelton MDF (MA3) (t-1)  Crakehill MDF (MA4) (t-1)  \
6                      NaN                       NaN

   Skip Bridge MDF (MA4) (t-1)  Westwick MDF (MA4) (t-1)  \
6                          NaN                       NaN

   Skelton MDF (MA4) (t-1)  …  Malham Tarn DRT (MA5) (t-1)  \
6                      NaN  …                        NaN

   Snaizeholme DRT (MA5) (t-1)  Arkengarthdale DRT (MA6) (t-1)  \
6                          NaN                           NaN

   East Cowton DRT (MA6) (t-1)  Malham Tarn DRT (MA6) (t-1)  \
6                          NaN                         NaN

   Snaizeholme DRT (MA6) (t-1)  Arkengarthdale DRT (MA7) (t-1)  \
6                          NaN                           NaN
```

```
   East Cowton DRT (MA7) (t-1)  Malham Tarn DRT (MA7) (t-1)  \
6                          NaN                          NaN

   Snaizeholme DRT (MA7) (t-1)
6                          NaN

[1 rows x 42 columns]
```

[93]:
```
lagged_ma_df.dropna(how="any", inplace=True)
lagged_ma_df[lagged_ma_df.isna().any(axis=1)]
```

[93]:
```
Empty DataFrame
Columns: [Date, Skelton MDF (Cumecs), Crakehill MDF (MA3) (t-1), Skip Bridge MDF
(MA3) (t-1), Westwick MDF (MA3) (t-1), Skelton MDF (MA3) (t-1), Crakehill MDF
(MA4) (t-1), Skip Bridge MDF (MA4) (t-1), Westwick MDF (MA4) (t-1), Skelton MDF
(MA4) (t-1), Crakehill MDF (MA5) (t-1), Skip Bridge MDF (MA5) (t-1), Westwick
MDF (MA5) (t-1), Skelton MDF (MA5) (t-1), Crakehill MDF (MA6) (t-1), Skip Bridge
MDF (MA6) (t-1), Westwick MDF (MA6) (t-1), Skelton MDF (MA6) (t-1), Crakehill
MDF (MA7) (t-1), Skip Bridge MDF (MA7) (t-1), Westwick MDF (MA7) (t-1), Skelton
MDF (MA7) (t-1), Arkengarthdale DRT (MA3) (t-1), East Cowton DRT (MA3) (t-1),
Malham Tarn DRT (MA3) (t-1), Snaizeholme DRT (MA3) (t-1), Arkengarthdale DRT
(MA4) (t-1), East Cowton DRT (MA4) (t-1), Malham Tarn DRT (MA4) (t-1),
Snaizeholme DRT (MA4) (t-1), Arkengarthdale DRT (MA5) (t-1), East Cowton DRT
(MA5) (t-1), Malham Tarn DRT (MA5) (t-1), Snaizeholme DRT (MA5) (t-1),
Arkengarthdale DRT (MA6) (t-1), East Cowton DRT (MA6) (t-1), Malham Tarn DRT
(MA6) (t-1), Snaizeholme DRT (MA6) (t-1), Arkengarthdale DRT (MA7) (t-1), East
Cowton DRT (MA7) (t-1), Malham Tarn DRT (MA7) (t-1), Snaizeholme DRT (MA7)
(t-1)]
Index: []

[0 rows x 42 columns]
```

[94]:
```
lagged_ma_df.describe()
```

[94]:
```
       Skelton MDF (Cumecs)  Crakehill MDF (MA3) (t-1)  \
count           1444.000000                1444.000000
mean              46.988251                  19.548435
std               55.796058                  23.168519
min                3.694000                   2.083333
25%               12.352500                   5.377500
50%               24.230000                  10.263333
75%               59.100000                  24.016667
max              448.100000                 176.666667

       Skip Bridge MDF (MA3) (t-1)  Westwick MDF (MA3) (t-1)  \
count                  1444.000000               1444.000000
```

mean	7.609933	21.521265
std	9.375087	25.111456
min	1.012667	1.990333
25%	2.174250	5.824167
50%	3.567667	10.922667
75%	9.570917	29.109167
max	67.550333	251.526667

	Skelton MDF (MA3) (t-1)	Crakehill MDF (MA4) (t-1)	\
count	1444.000000	1444.000000	
mean	47.037087	19.550722	
std	52.773360	22.406724	
min	3.733000	2.095000	
25%	12.803333	5.574375	
50%	24.756667	10.430000	
75%	62.124167	25.050000	
max	361.066667	158.250000	

	Skip Bridge MDF (MA4) (t-1)	Westwick MDF (MA4) (t-1)	\
count	1444.000000	1444.000000	
mean	7.610739	21.527241	
std	9.119847	24.219128	
min	1.014750	2.003750	
25%	2.178812	5.848125	
50%	3.592375	11.259750	
75%	9.585000	29.589500	
max	65.921750	216.440250	

	Skelton MDF (MA4) (t-1)	Crakehill MDF (MA5) (t-1)	…	\
count	1444.000000	1444.000000	…	
mean	47.043488	19.551895	…	
std	51.547477	21.775279	…	
min	3.834250	2.110000	…	
25%	13.049875	5.694000	…	
50%	25.341250	10.803000	…	
75%	63.085000	25.390000	…	
max	337.300000	157.960000	…	

	Malham Tarn DRT (MA5) (t-1)	Snaizeholme DRT (MA5) (t-1)	\
count	1444.000000	1444.000000	
mean	14.033352	9.618393	
std	15.173028	12.912583	
min	0.000000	0.000000	
25%	1.760000	0.800000	
50%	9.760000	4.960000	
75%	21.140000	13.300000	
max	90.240000	105.040000	

```
       Arkengarthdale DRT (MA6) (t-1)  East Cowton DRT (MA6) (t-1)  \
count                     1444.000000                  1444.000000
mean                         5.792336                     5.291043
std                          7.303841                     6.080526
min                          0.000000                     0.000000
25%                          0.666667                     0.933333
50%                          3.466667                     3.466667
75%                          8.400000                     7.266667
max                         73.400000                    40.400000

       Malham Tarn DRT (MA6) (t-1)  Snaizeholme DRT (MA6) (t-1)  \
count                 1.444000e+03                  1444.000000
mean                  1.403389e+01                     9.627054
std                   1.446474e+01                    12.377981
min                   8.659740e-15                     0.000000
25%                   2.133333e+00                     1.200000
50%                   1.033333e+01                     5.333333
75%                   2.140000e+01                    13.333333
max                   9.733333e+01                    87.800000

       Arkengarthdale DRT (MA7) (t-1)  East Cowton DRT (MA7) (t-1)  \
count                    1.444000e+03                  1444.000000
mean                     5.791848e+00                     5.290305
std                      6.985875e+00                     5.733811
min                      5.075305e-16                     0.000000
25%                      8.000000e-01                     1.257143
50%                      3.771429e+00                     3.657143
75%                      8.342857e+00                     7.100000
max                      6.965714e+01                    35.200000

       Malham Tarn DRT (MA7) (t-1)  Snaizeholme DRT (MA7) (t-1)
count                 1.444000e+03                  1444.000000
mean                  1.403419e+01                     9.633478
std                   1.387652e+01                    11.913562
min                   5.773160e-15                     0.000000
25%                   2.171429e+00                     1.371429
50%                   1.051429e+01                     5.828571
75%                   2.104286e+01                    13.371429
max                   1.040571e+02                    75.257143

[8 rows x 41 columns]
```

[95]:
```python
# Creating weighted moving averages
weighted_ma_df = pd.DataFrame()
weighted_ma_df["Date"] = clean_df["Date"]
weighted_ma_df[flow_cols[-1]] = clean_df[flow_cols[-1]]
```

```
weighted_ma_df.head(10)
```

[95]:

	Date	Skelton MDF (Cumecs)
0	1993-01-01	26.10
1	1993-01-02	24.86
2	1993-01-03	23.60
3	1993-01-04	23.47
4	1993-01-05	60.70
5	1993-01-06	98.01
6	1993-01-07	56.99
7	1993-01-08	56.66
8	1993-01-09	78.10
9	1993-01-10	125.70

[96]:
```
## Creating weighted moving averages of between 3 and 7 days for each numerical column
for i in range(3, 8):
    for col in flow_cols:
        col_name = col.replace("(Cumecs)", f"(WMA{i})")
        weighted_ma_df[col_name] = clean_df[col].ewm(span=i).mean()

for i in range(3, 8):
    for col in rain_cols:
        col_name = col.replace("(mm)", f"(WMA{i})")
        weighted_ma_df[col_name] = clean_df[col].ewm(span=i).mean()

weighted_ma_df.head(10)
```

[96]:

	Date	Skelton MDF (Cumecs)	Crakehill MDF (WMA3)	\
0	1993-01-01	26.10	10.400000	
1	1993-01-02	24.86	10.100000	
2	1993-01-03	23.60	9.734286	
3	1993-01-04	23.47	9.561333	
4	1993-01-05	60.70	18.200645	
5	1993-01-06	98.01	25.260635	
6	1993-01-07	56.99	22.256850	
7	1993-01-08	56.66	22.127922	
8	1993-01-09	78.10	28.827045	
9	1993-01-10	125.70	39.924360	

	Skip Bridge MDF (WMA3)	Westwick MDF (WMA3)	Skelton MDF (WMA3)	\
0	4.393000	9.291000	26.100000	
1	4.290333	8.845000	25.273333	
2	4.195286	8.394714	24.317143	
3	4.284733	8.144200	23.865333	
4	8.247194	34.239581	42.876774	
5	9.257889	34.329190	70.880952	

```
6                8.248055            28.248591           63.880787
7                7.755102            28.920420           60.256235
8                7.954440            44.617368           69.195577
9               10.617821            69.308296           97.475406

   Crakehill MDF (WMA4)  Skip Bridge MDF (WMA4)  Westwick MDF (WMA4)  \
0             10.400000                4.393000             9.291000
1             10.118750                4.296750             8.872875
2              9.782653                4.208612             8.456612
3              9.611397                4.279562             8.212305
4             16.849688                7.611640            30.111895
5             23.248349                8.713177            31.917793
6             21.623524                8.112697            27.944653
7             21.776687                7.768232            28.612815
8             27.321895                7.923706            41.397731
9             36.850755               10.077648            62.546920

   Skelton MDF (WMA4)  …  Malham Tarn DRT (WMA5)  Snaizeholme DRT (WMA5)  \
0           26.100000  …                0.000000                4.000000
1           25.325000  …                0.480000                1.600000
2           24.444898  …                0.631579                0.842105
3           23.996875  …                0.701538               26.080000
4           39.915996  …               13.330806               58.756398
5           64.290830  …                9.044211               37.578346
6           61.286393  …               12.073434               37.019524
7           59.404222  …                8.440539               25.011134
8           66.958665  …               24.443295               52.317730
9           90.598138  …               41.932146               80.975455

   Arkengarthdale DRT (WMA6)  East Cowton DRT (WMA6)  Malham Tarn DRT (WMA6)  \
0                   0.000000                0.000000                0.000000
1                   0.000000                0.000000                0.466667
2                   0.000000                0.000000                0.616514
3                   0.927027                9.579279                0.687387
4                   4.532554                8.182663               12.238795
5                   3.039206                5.486709                8.733616
6                   3.847544                5.017416               11.532371
7                   3.158713                3.479670                8.488282
8                   6.533868                2.675109               22.513302
9                  10.755867                2.593692               38.342455

   Snaizeholme DRT (WMA6)  Arkengarthdale DRT (WMA7)  East Cowton DRT (WMA7)  \
0                4.000000                   0.000000                0.000000
1                1.666667                   0.000000                0.000000
2                0.917431                   0.000000                0.000000
3               24.356757                   0.877714                9.069714
4               54.836310                   4.261204                7.932394
```

```
5               37.032886                   2.965251                5.519929
6               36.706846                   3.725407                5.081412
7               26.192430                   3.134942                3.669731
8               49.794060                   6.179831                2.894057
9               75.543086                  10.052984                2.763172

   Malham Tarn DRT (WMA7)  Snaizeholme DRT (WMA7)
0                0.000000                4.000000
1                0.457143                1.714286
2                0.605405                0.972973
3                0.676571               23.145143
4               11.468374               52.008195
5                8.467122               36.434333
6               11.102064               36.309023
7                8.462270               26.888668
8               21.095246               47.947713
9               35.640530               71.486327

[10 rows x 42 columns]
```

[97]:
```
weighted_ma_df.describe()
```

[97]:
```
       Skelton MDF (Cumecs)  Crakehill MDF (WMA3)  Skip Bridge MDF (WMA3)  \
count           1451.000000           1451.000000             1451.000000
mean              46.977784             19.519175                7.600024
std               55.691669             22.509756                9.131015
min                3.694000              2.106613                1.013019
25%               12.385000              5.537138                2.189022
50%               24.250000             10.408988                3.595440
75%               59.150000             24.015073                9.343599
max              448.100000            171.548253               66.767533

       Westwick MDF (WMA3)  Skelton MDF (WMA3)  Crakehill MDF (WMA4)  \
count          1451.000000         1451.000000           1451.000000
mean             21.492349           46.971894             19.518767
std              24.494882           51.546413             21.708574
min               1.988822            3.784612              2.124633
25%               5.931549           12.991008              5.664710
50%              11.193894           25.273333             10.764707
75%              29.361253           63.598434             24.935907
max             283.847215          358.363081            163.374574

       Skip Bridge MDF (WMA4)  Westwick MDF (WMA4)  Skelton MDF (WMA4)  \
count             1451.000000          1451.000000         1451.000000
mean                 7.599538            21.492851           46.971182
std                  8.854120            23.552924           50.203609
min                  1.015611             2.003615            3.813018
```

```
25%                  2.200730            6.034924           13.232278
50%                  3.623453           11.656633           25.870124
75%                  9.732283           29.933482           62.757663
max                 63.330198          252.563531          342.371380

       Crakehill MDF (WMA5)  …  Malham Tarn DRT (WMA5)  \
count           1451.000000  …             1451.000000
mean              19.518829  …               13.984398
std               21.070805  …               14.834316
min                2.143048  …                0.000000
25%                5.781697  …                2.396532
50%               11.025246  …                9.522856
75%               24.765063  …               21.230186
max              155.556128  …              105.652838

       Snaizeholme DRT (WMA5)  Arkengarthdale DRT (WMA6)  \
count             1451.000000                1451.000000
mean                 9.651831                   5.775847
std                 12.682914                   7.117200
min                  0.000023                   0.000000
25%                  1.361718                   1.034738
50%                  5.152346                   3.629764
75%                 12.726148                   8.114486
max                 97.166216                  83.577971

       East Cowton DRT (WMA6)  Malham Tarn DRT (WMA6)  Snaizeholme DRT (WMA6)  \
count             1451.000000             1451.000000             1451.000000
mean                 5.286105               13.985534                9.659651
std                  5.928408               14.046123               12.084998
min                  0.000000                0.000000                0.000130
25%                  1.297775                2.864697                1.573018
50%                  3.479670               10.051803                5.627342
75%                  7.154646               21.351626               12.777683
max                 51.966182              100.117414               86.247750

       Arkengarthdale DRT (WMA7)  East Cowton DRT (WMA7)  \
count                1451.000000             1451.000000
mean                    5.775058                5.284842
std                     6.759876                5.554247
min                     0.000000                0.000000
25%                     1.227319                1.504496
50%                     3.795612                3.669731
75%                     8.044823                7.163909
max                    77.825716               45.744930

       Malham Tarn DRT (WMA7)  Snaizeholme DRT (WMA7)
count             1451.000000             1451.000000
```

```
mean                13.986942                9.668315
std                 13.408911               11.612662
min                  0.000000                0.000424
25%                  3.248877                1.729078
50%                 10.562586                5.856426
75%                 21.049487               12.844838
max                 95.347821               80.502521

[8 rows x 41 columns]
```

[98]:
```
# Creating lagged weighted moving averages
lagged_wma_df = pd.DataFrame()
lagged_wma_df["Date"] = weighted_ma_df["Date"]
lagged_wma_df[flow_cols[-1]] = weighted_ma_df[flow_cols[-1]]
lagged_wma_df.head(10)
```

[98]:

	Date	Skelton MDF (Cumecs)
0	1993-01-01	26.10
1	1993-01-02	24.86
2	1993-01-03	23.60
3	1993-01-04	23.47
4	1993-01-05	60.70
5	1993-01-06	98.01
6	1993-01-07	56.99
7	1993-01-08	56.66
8	1993-01-09	78.10
9	1993-01-10	125.70

[99]:
```
# Lagging weighted moving averages
w_mdf_cols = list(weighted_ma_df.columns[2:22])
w_drt_cols = list(weighted_ma_df.columns[22:])

for col in w_mdf_cols:
    col_name = col + f" (t-1)"
    lagged_wma_df[col_name] = weighted_ma_df[col].shift(1)

for col in w_drt_cols:
    col_name = col + f" (t-1)"
    lagged_wma_df[col_name] = weighted_ma_df[col].shift(1)

lagged_wma_df.head(10)
```

[99]:

	Date	Skelton MDF (Cumecs)	Crakehill MDF (WMA3) (t-1)	\
0	1993-01-01	26.10	NaN	
1	1993-01-02	24.86	10.400000	
2	1993-01-03	23.60	10.100000	
3	1993-01-04	23.47	9.734286	

```
4  1993-01-05                60.70                    9.561333
5  1993-01-06                98.01                   18.200645
6  1993-01-07                56.99                   25.260635
7  1993-01-08                56.66                   22.256850
8  1993-01-09                78.10                   22.127922
9  1993-01-10               125.70                   28.827045

   Skip Bridge MDF (WMA3) (t-1)  Westwick MDF (WMA3) (t-1)  \
0                           NaN                         NaN
1                      4.393000                    9.291000
2                      4.290333                    8.845000
3                      4.195286                    8.394714
4                      4.284733                    8.144200
5                      8.247194                   34.239581
6                      9.257889                   34.329190
7                      8.248055                   28.248591
8                      7.755102                   28.920420
9                      7.954440                   44.617368

   Skelton MDF (WMA3) (t-1)  Crakehill MDF (WMA4) (t-1)  \
0                       NaN                         NaN
1                 26.100000                   10.400000
2                 25.273333                   10.118750
3                 24.317143                    9.782653
4                 23.865333                    9.611397
5                 42.876774                   16.849688
6                 70.880952                   23.248349
7                 63.880787                   21.623524
8                 60.256235                   21.776687
9                 69.195577                   27.321895

   Skip Bridge MDF (WMA4) (t-1)  Westwick MDF (WMA4) (t-1)  \
0                           NaN                         NaN
1                      4.393000                    9.291000
2                      4.296750                    8.872875
3                      4.208612                    8.456612
4                      4.279562                    8.212305
5                      7.611640                   30.111895
6                      8.713177                   31.917793
7                      8.112697                   27.944653
8                      7.768232                   28.612815
9                      7.923706                   41.397731

   Skelton MDF (WMA4) (t-1)  …  Malham Tarn DRT (WMA5) (t-1)  \
0                       NaN  …                           NaN
1                 26.100000  …                      0.000000
2                 25.325000  …                      0.480000
```

```
3                24.444898  …                    0.631579
4                23.996875  …                    0.701538
5                39.915996  …                   13.330806
6                64.290830  …                    9.044211
7                61.286393  …                   12.073434
8                59.404222  …                    8.440539
9                66.958665  …                   24.443295

   Snaizeholme DRT (WMA5) (t-1)  Arkengarthdale DRT (WMA6) (t-1)  \
0                           NaN                           NaN
1                      4.000000                      0.000000
2                      1.600000                      0.000000
3                      0.842105                      0.000000
4                     26.080000                      0.927027
5                     58.756398                      4.532554
6                     37.578346                      3.039206
7                     37.019524                      3.847544
8                     25.011134                      3.158713
9                     52.317730                      6.533868

   East Cowton DRT (WMA6) (t-1)  Malham Tarn DRT (WMA6) (t-1)  \
0                           NaN                           NaN
1                      0.000000                      0.000000
2                      0.000000                      0.466667
3                      0.000000                      0.616514
4                      9.579279                      0.687387
5                      8.182663                     12.238795
6                      5.486709                      8.733616
7                      5.017416                     11.532371
8                      3.479670                      8.488282
9                      2.675109                     22.513302

   Snaizeholme DRT (WMA6) (t-1)  Arkengarthdale DRT (WMA7) (t-1)  \
0                           NaN                           NaN
1                      4.000000                      0.000000
2                      1.666667                      0.000000
3                      0.917431                      0.000000
4                     24.356757                      0.877714
5                     54.836310                      4.261204
6                     37.032886                      2.965251
7                     36.706846                      3.725407
8                     26.192430                      3.134942
9                     49.794060                      6.179831

   East Cowton DRT (WMA7) (t-1)  Malham Tarn DRT (WMA7) (t-1)  \
0                           NaN                           NaN
1                      0.000000                      0.000000
```

```
2                      0.000000                      0.457143
3                      0.000000                      0.605405
4                      9.069714                      0.676571
5                      7.932394                     11.468374
6                      5.519929                      8.467122
7                      5.081412                     11.102064
8                      3.669731                      8.462270
9                      2.894057                     21.095246

   Snaizeholme DRT (WMA7) (t-1)
0                           NaN
1                      4.000000
2                      1.714286
3                      0.972973
4                     23.145143
5                     52.008195
6                     36.434333
7                     36.309023
8                     26.888668
9                     47.947713

[10 rows x 42 columns]
```

[100]:
```
lagged_wma_df[lagged_wma_df.isna().any(1)]
```

[100]:
```
         Date  Skelton MDF (Cumecs)  Crakehill MDF (WMA3) (t-1)  \
0  1993-01-01                  26.1                         NaN

   Skip Bridge MDF (WMA3) (t-1)  Westwick MDF (WMA3) (t-1)  \
0                           NaN                        NaN

   Skelton MDF (WMA3) (t-1)  Crakehill MDF (WMA4) (t-1)  \
0                       NaN                         NaN

   Skip Bridge MDF (WMA4) (t-1)  Westwick MDF (WMA4) (t-1)  \
0                           NaN                        NaN

   Skelton MDF (WMA4) (t-1)  …  Malham Tarn DRT (WMA5) (t-1)  \
0                       NaN  …                          NaN

   Snaizeholme DRT (WMA5) (t-1)  Arkengarthdale DRT (WMA6) (t-1)  \
0                           NaN                             NaN

   East Cowton DRT (WMA6) (t-1)  Malham Tarn DRT (WMA6) (t-1)  \
0                           NaN                          NaN

   Snaizeholme DRT (WMA6) (t-1)  Arkengarthdale DRT (WMA7) (t-1)  \
```

```
0                           NaN                                  NaN

   East Cowton DRT (WMA7) (t-1)  Malham Tarn DRT (WMA7) (t-1)  \
0                           NaN                           NaN

   Snaizeholme DRT (WMA7) (t-1)
0                           NaN

[1 rows x 42 columns]
```

[101]:
```
lagged_wma_df.dropna(how="any", inplace=True)
lagged_wma_df[lagged_wma_df.isna().any(1)]
```

[101]:
```
Empty DataFrame
Columns: [Date, Skelton MDF (Cumecs), Crakehill MDF (WMA3) (t-1), Skip Bridge
MDF (WMA3) (t-1), Westwick MDF (WMA3) (t-1), Skelton MDF (WMA3) (t-1), Crakehill
MDF (WMA4) (t-1), Skip Bridge MDF (WMA4) (t-1), Westwick MDF (WMA4) (t-1),
Skelton MDF (WMA4) (t-1), Crakehill MDF (WMA5) (t-1), Skip Bridge MDF (WMA5)
(t-1), Westwick MDF (WMA5) (t-1), Skelton MDF (WMA5) (t-1), Crakehill MDF (WMA6)
(t-1), Skip Bridge MDF (WMA6) (t-1), Westwick MDF (WMA6) (t-1), Skelton MDF
(WMA6) (t-1), Crakehill MDF (WMA7) (t-1), Skip Bridge MDF (WMA7) (t-1), Westwick
MDF (WMA7) (t-1), Skelton MDF (WMA7) (t-1), Arkengarthdale DRT (WMA3) (t-1),
East Cowton DRT (WMA3) (t-1), Malham Tarn DRT (WMA3) (t-1), Snaizeholme DRT
(WMA3) (t-1), Arkengarthdale DRT (WMA4) (t-1), East Cowton DRT (WMA4) (t-1),
Malham Tarn DRT (WMA4) (t-1), Snaizeholme DRT (WMA4) (t-1), Arkengarthdale DRT
(WMA5) (t-1), East Cowton DRT (WMA5) (t-1), Malham Tarn DRT (WMA5) (t-1),
Snaizeholme DRT (WMA5) (t-1), Arkengarthdale DRT (WMA6) (t-1), East Cowton DRT
(WMA6) (t-1), Malham Tarn DRT (WMA6) (t-1), Snaizeholme DRT (WMA6) (t-1),
Arkengarthdale DRT (WMA7) (t-1), East Cowton DRT (WMA7) (t-1), Malham Tarn DRT
(WMA7) (t-1), Snaizeholme DRT (WMA7) (t-1)]
Index: []

[0 rows x 42 columns]
```

[102]:
```
lagged_wma_df.describe()
```

[102]:
```
       Skelton MDF (Cumecs)  Crakehill MDF (WMA3) (t-1)  \
count           1450.000000                 1450.000000
mean              46.992182                   19.522053
std               55.708181                   22.517255
min                3.694000                    2.106613
25%               12.382500                    5.534470
50%               24.245000                   10.407141
75%               59.200000                   24.051890
max              448.100000                  171.548253

       Skip Bridge MDF (WMA3) (t-1)  Westwick MDF (WMA3) (t-1)  \
```

```
       Skelton MDF (WMA3) (t-1)  Crakehill MDF (WMA4) (t-1)  \
count               1450.000000                 1450.000000
mean                  46.979458                   19.521839
std                   51.563392                   21.715748
min                    3.784612                    2.124633
25%                   12.988892                    5.652475
50%                   25.272767                   10.740245
75%                   63.638607                   24.975226
max                  358.363081                  163.374574

       Skip Bridge MDF (WMA4) (t-1)  Westwick MDF (WMA4) (t-1)  \
count                   1450.000000                 1450.000000
mean                       7.600470                   21.498375
std                        8.857104                   23.560109
min                        1.015611                    2.003615
25%                        2.199878                    6.030254
50%                        3.621673                   11.629449
75%                        9.753887                   29.937759
max                       63.330198                  252.563531

       Skelton MDF (WMA4) (t-1)  Crakehill MDF (WMA5) (t-1)  …  \
count               1450.000000                 1450.000000  …
mean                  46.978713                   19.521840  …
std                   50.220109                   21.077762  …
min                    3.813018                    2.143048  …
25%                   13.230015                    5.778349  …
50%                   25.859974                   11.015578  …
75%                   62.787376                   24.787001  …
max                  342.371380                  155.556128  …

       Malham Tarn DRT (WMA5) (t-1)  Snaizeholme DRT (WMA5) (t-1)  \
count                   1450.000000                   1450.000000
mean                      13.990728                      9.656912
std                       14.837473                     12.685812
min                        0.000000                      0.000023
25%                        2.392643                      1.360390
50%                        9.528469                      5.157073
75%                       21.253441                     12.729215
```

```
max                      105.652838                    97.166216

       Arkengarthdale DRT (WMA6) (t-1)  East Cowton DRT (WMA6) (t-1)  \
count                      1450.000000                   1450.000000
mean                          5.775426                      5.284441
std                           7.119637                      5.930115
min                           0.000000                      0.000000
25%                           1.033939                      1.293475
50%                           3.628686                      3.478350
75%                           8.118800                      7.146410
max                          83.577971                     51.966182

       Malham Tarn DRT (WMA6) (t-1)  Snaizeholme DRT (WMA6) (t-1)  \
count                   1450.000000                   1450.000000
mean                      13.991815                      9.664587
std                       14.048930                     12.087704
min                        0.000000                      0.000130
25%                        2.855575                      1.571982
50%                       10.064458                      5.628753
75%                       21.354523                     12.786220
max                      100.117414                     86.247750

       Arkengarthdale DRT (WMA7) (t-1)  East Cowton DRT (WMA7) (t-1)  \
count                      1450.000000                   1450.000000
mean                          5.774654                      5.283198
std                           6.762191                      5.555810
min                           0.000000                      0.000000
25%                           1.227186                      1.504081
50%                           3.792445                      3.664996
75%                           8.049764                      7.160675
max                          77.825716                     45.744930

       Malham Tarn DRT (WMA7) (t-1)  Snaizeholme DRT (WMA7) (t-1)
count                   1450.000000                   1450.000000
mean                      13.993078                      9.673086
std                       13.411499                     11.615246
min                        0.000000                      0.000424
25%                        3.244706                      1.728840
50%                       10.593781                      5.863176
75%                       21.058371                     12.848060
max                       95.347821                     80.502521

[8 rows x 41 columns]
```

2.0.3 Exporting Datasets

[103]:
```
lagged_df.describe()
```

[103]:

	Skelton MDF (Cumecs)	Crakehill MDF (t-1)	Skip Bridge MDF (t-1)	\
count	1448.000000	1448.000000	1448.000000	
mean	47.023622	19.537023	7.606707	
std	55.740230	25.272741	10.023280	
min	3.694000	2.060000	1.002000	
25%	12.377500	5.215000	2.154500	
50%	24.245000	10.050000	3.480500	
75%	59.270000	21.800000	8.855250	
max	448.100000	220.000000	80.244000	

	Westwick MDF (t-1)	Skelton MDF (t-1)	Crakehill MDF (t-2)	\
count	1448.000000	1448.000000	1448.000000	
mean	21.517077	47.012696	19.532983	
std	28.287685	55.743271	25.273807	
min	1.954000	3.694000	2.060000	
25%	5.609250	12.377500	5.215000	
50%	10.385000	24.230000	10.000000	
75%	26.724000	59.270000	21.800000	
max	374.061000	448.100000	220.000000	

	Skip Bridge MDF (t-2)	Westwick MDF (t-2)	Skelton MDF (t-2)	\
count	1448.000000	1448.000000	1448.000000	
mean	7.605369	21.513186	47.006191	
std	10.023601	28.289072	55.745308	
min	1.002000	1.954000	3.694000	
25%	2.154500	5.609250	12.377500	
50%	3.480500	10.354500	24.230000	
75%	8.855250	26.724000	59.270000	
max	80.244000	374.061000	448.100000	

	Crakehill MDF (t-3)	…	Malham Tarn DRT (t-1)	Snaizeholme DRT (t-1)	\
count	1448.000000	…	1448.000000	1448.000000	
mean	19.531809	…	14.010497	9.652486	
std	25.274192	…	25.065902	20.953300	
min	2.060000	…	0.000000	0.000000	
25%	5.215000	…	0.000000	0.000000	
50%	10.000000	…	1.600000	0.800000	
75%	21.800000	…	18.400000	10.400000	
max	220.000000	…	252.000000	268.800000	

	Arkengarthdale DRT (t-2)	East Cowton DRT (t-2)	Malham Tarn DRT (t-2)	\
count	1448.000000	1448.000000	1448.000000	
mean	5.783978	5.295028	14.007182	

```
std                   13.274272             12.308373             25.067332
min                    0.000000              0.000000              0.000000
25%                    0.000000              0.000000              0.000000
50%                    0.800000              0.000000              1.600000
75%                    6.400000              5.300000             18.400000
max                  225.200000            165.600000            252.000000

       Snaizeholme DRT (t-2)  Arkengarthdale DRT (t-3)  East Cowton DRT (t-3)  \
count            1448.000000               1448.000000            1448.000000
mean                9.649724                  5.781768               5.291713
std                20.954309                 13.274969              12.309152
min                 0.000000                  0.000000               0.000000
25%                 0.000000                  0.000000               0.000000
50%                 0.800000                  0.800000               0.000000
75%                10.400000                  6.400000               5.300000
max               268.800000                225.200000             165.600000

       Malham Tarn DRT (t-3)  Snaizeholme DRT (t-3)
count            1448.000000            1448.000000
mean               14.007182               9.651934
std                25.067332              20.953544
min                 0.000000               0.000000
25%                 0.000000               0.000000
50%                 1.600000               0.800000
75%                18.400000              10.400000
max               252.000000             268.800000

[8 rows x 25 columns]
```

[104]:
```
moving_avg_df.describe()
```

[104]:
```
       Skelton MDF (Cumecs)  Crakehill MDF (MA3)  Skip Bridge MDF (MA3)  \
count           1445.000000          1445.000000            1445.000000
mean              46.995172            19.545218               7.608666
std               55.777355            23.160818               9.371964
min                3.694000             2.083333               1.012667
25%               12.370000             5.383333               2.176333
50%               24.240000            10.276667               3.568667
75%               59.050000            23.966667               9.565000
max              448.100000           176.666667              67.550333

       Westwick MDF (MA3)  Skelton MDF (MA3)  Crakehill MDF (MA4)  \
count         1445.000000        1445.000000          1445.000000
mean            21.515503          47.028298            19.546846
std             25.103715          52.756141            22.399448
min              1.990333           3.733000             2.095000
25%              5.825667          12.820000             5.585000
```

```
50%            10.934333         24.773333           10.435000
75%            29.075333         62.006667           25.050000
max           251.526667        361.066667          158.250000

       Skip Bridge MDF (MA4)  Westwick MDF (MA4)  Skelton MDF (MA4)  \
count            1445.000000         1445.000000        1445.000000
mean                7.609420           21.521166          47.033714
std                 9.116826           24.211841          51.530965
min                 1.014750            2.003750           3.834250
25%                 2.179500            5.850000          13.050000
50%                 3.593250           11.279000          25.365000
75%                 9.582250           29.579000          63.012500
max                65.921750          216.440250         337.300000

       Crakehill MDF (MA5)  …  Malham Tarn DRT (MA5)  Snaizeholme DRT (MA5)  \
count          1445.000000  …            1445.000000            1445.000000
mean             19.547596  …              14.027183               9.613619
std              21.768351  …              15.169586              12.909387
min               2.110000  …               0.000000               0.000000
25%               5.696000  …               1.760000               0.800000
50%              10.806000  …               9.760000               4.960000
75%              25.380000  …              21.120000              13.280000
max             157.960000  …              90.240000             105.040000

       Arkengarthdale DRT (MA6)  East Cowton DRT (MA6)  Malham Tarn DRT (MA6)  \
count                1445.000000            1445.000000           1.445000e+03
mean                    5.792203               5.292826           1.402713e+01
std                     7.301313               6.078797           1.446201e+01
min                     0.000000               0.000000           8.659740e-15
25%                     0.666667               0.933333           2.133333e+00
50%                     3.466667               3.466667           1.033333e+01
75%                     8.400000               7.266667           2.140000e+01
max                    73.400000              40.400000           9.733333e+01

       Snaizeholme DRT (MA6)  Arkengarthdale DRT (MA7)  East Cowton DRT (MA7)  \
count            1445.000000              1.445000e+03            1445.000000
mean                9.621961              5.791162e+00               5.291626
std                12.375209              6.983505e+00               5.732045
min                 0.000000              5.075305e-16               0.000000
25%                 1.200000              8.000000e-01               1.257143
50%                 5.333333              3.771429e+00               3.657143
75%                13.333333              8.342857e+00               7.142857
max                87.800000              6.965714e+01              35.200000

       Malham Tarn DRT (MA7)  Snaizeholme DRT (MA7)
count           1.445000e+03            1445.000000
mean            1.402709e+01               9.628314
```

```
std             1.387434e+01              11.911054
min             5.773160e-15               0.000000
25%             2.171429e+00               1.371429
50%             1.051429e+01               5.828571
75%             2.102857e+01              13.371429
max             1.040571e+02              75.257143

[8 rows x 41 columns]
```

[105]:
```
lagged_ma_df.describe()
```

[105]:
```
       Skelton MDF (Cumecs)  Crakehill MDF (MA3) (t-1)  \
count           1444.000000                1444.000000
mean              46.988251                  19.548435
std               55.796058                  23.168519
min                3.694000                   2.083333
25%               12.352500                   5.377500
50%               24.230000                  10.263333
75%               59.100000                  24.016667
max              448.100000                 176.666667

       Skip Bridge MDF (MA3) (t-1)  Westwick MDF (MA3) (t-1)  \
count                  1444.000000               1444.000000
mean                      7.609933                 21.521265
std                       9.375087                 25.111456
min                       1.012667                  1.990333
25%                       2.174250                  5.824167
50%                       3.567667                 10.922667
75%                       9.570917                 29.109167
max                      67.550333                251.526667

       Skelton MDF (MA3) (t-1)  Crakehill MDF (MA4) (t-1)  \
count              1444.000000                1444.000000
mean                 47.037087                  19.550722
std                  52.773360                  22.406724
min                   3.733000                   2.095000
25%                  12.803333                   5.574375
50%                  24.756667                  10.430000
75%                  62.124167                  25.050000
max                 361.066667                 158.250000

       Skip Bridge MDF (MA4) (t-1)  Westwick MDF (MA4) (t-1)  \
count                  1444.000000               1444.000000
mean                      7.610739                 21.527241
std                       9.119847                 24.219128
min                       1.014750                  2.003750
25%                       2.178812                  5.848125
```

```
50%                      3.592375                 11.259750
75%                      9.585000                 29.589500
max                     65.921750                216.440250

       Skelton MDF (MA4) (t-1)  Crakehill MDF (MA5) (t-1)  …  \
count              1444.000000                1444.000000  …
mean                 47.043488                  19.551895  …
std                  51.547477                  21.775279  …
min                   3.834250                   2.110000  …
25%                  13.049875                   5.694000  …
50%                  25.341250                  10.803000  …
75%                  63.085000                  25.390000  …
max                 337.300000                 157.960000  …

       Malham Tarn DRT (MA5) (t-1)  Snaizeholme DRT (MA5) (t-1)  \
count                  1444.000000                  1444.000000
mean                     14.033352                     9.618393
std                      15.173028                    12.912583
min                       0.000000                     0.000000
25%                       1.760000                     0.800000
50%                       9.760000                     4.960000
75%                      21.140000                    13.300000
max                      90.240000                   105.040000

       Arkengarthdale DRT (MA6) (t-1)  East Cowton DRT (MA6) (t-1)  \
count                     1444.000000                  1444.000000
mean                        5.792336                     5.291043
std                         7.303841                     6.080526
min                         0.000000                     0.000000
25%                         0.666667                     0.933333
50%                         3.466667                     3.466667
75%                         8.400000                     7.266667
max                        73.400000                    40.400000

       Malham Tarn DRT (MA6) (t-1)  Snaizeholme DRT (MA6) (t-1)  \
count                 1.444000e+03                  1444.000000
mean                  1.403389e+01                     9.627054
std                   1.446474e+01                    12.377981
min                   8.659740e-15                     0.000000
25%                   2.133333e+00                     1.200000
50%                   1.033333e+01                     5.333333
75%                   2.140000e+01                    13.333333
max                   9.733333e+01                    87.800000

       Arkengarthdale DRT (MA7) (t-1)  East Cowton DRT (MA7) (t-1)  \
count                     1.444000e+03                  1444.000000
mean                      5.791848e+00                     5.290305
```

```
std                     6.985875e+00                    5.733811
min                     5.075305e-16                    0.000000
25%                     8.000000e-01                    1.257143
50%                     3.771429e+00                    3.657143
75%                     8.342857e+00                    7.100000
max                     6.965714e+01                   35.200000

       Malham Tarn DRT (MA7) (t-1)  Snaizeholme DRT (MA7) (t-1)
count                 1.444000e+03                  1444.000000
mean                  1.403419e+01                     9.633478
std                   1.387652e+01                    11.913562
min                   5.773160e-15                     0.000000
25%                   2.171429e+00                     1.371429
50%                   1.051429e+01                     5.828571
75%                   2.104286e+01                    13.371429
max                   1.040571e+02                    75.257143

[8 rows x 41 columns]
```

[106]:
```
lagged_wma_df.describe()
```

[106]:
```
       Skelton MDF (Cumecs)  Crakehill MDF (WMA3) (t-1)  \
count           1450.000000                 1450.000000
mean              46.992182                   19.522053
std               55.708181                   22.517255
min                3.694000                    2.106613
25%               12.382500                    5.534470
50%               24.245000                   10.407141
75%               59.200000                   24.051890
max              448.100000                  171.548253

       Skip Bridge MDF (WMA3) (t-1)  Westwick MDF (WMA3) (t-1)  \
count                   1450.000000                1450.000000
mean                       7.601166                  21.497952
std                        9.134062                  24.502403
min                        1.013019                   1.988822
25%                        2.188817                   5.928907
50%                        3.595122                  11.170101
75%                        9.386454                  29.367607
max                       66.767533                 283.847215

       Skelton MDF (WMA3) (t-1)  Crakehill MDF (WMA4) (t-1)  \
count                1450.000000                 1450.000000
mean                   46.979458                   19.521839
std                    51.563392                   21.715748
min                     3.784612                    2.124633
25%                    12.988892                    5.652475
```

```
50%                   25.272767                   10.740245
75%                   63.638607                   24.975226
max                  358.363081                  163.374574

       Skip Bridge MDF (WMA4) (t-1)  Westwick MDF (WMA4) (t-1)  \
count                   1450.000000                 1450.000000
mean                       7.600470                   21.498375
std                        8.857104                   23.560109
min                        1.015611                    2.003615
25%                        2.199878                    6.030254
50%                        3.621673                   11.629449
75%                        9.753887                   29.937759
max                       63.330198                  252.563531

       Skelton MDF (WMA4) (t-1)  Crakehill MDF (WMA5) (t-1)  …  \
count               1450.000000                 1450.000000  …
mean                  46.978713                   19.521840  …
std                   50.220109                   21.077762  …
min                    3.813018                    2.143048  …
25%                   13.230015                    5.778349  …
50%                   25.859974                   11.015578  …
75%                   62.787376                   24.787001  …
max                  342.371380                  155.556128  …

       Malham Tarn DRT (WMA5) (t-1)  Snaizeholme DRT (WMA5) (t-1)  \
count                   1450.000000                   1450.000000
mean                      13.990728                      9.656912
std                       14.837473                     12.685812
min                        0.000000                      0.000023
25%                        2.392643                      1.360390
50%                        9.528469                      5.157073
75%                       21.253441                     12.729215
max                      105.652838                     97.166216

       Arkengarthdale DRT (WMA6) (t-1)  East Cowton DRT (WMA6) (t-1)  \
count                      1450.000000                   1450.000000
mean                          5.775426                      5.284441
std                           7.119637                      5.930115
min                           0.000000                      0.000000
25%                           1.033939                      1.293475
50%                           3.628686                      3.478350
75%                           8.118800                      7.146410
max                          83.577971                     51.966182

       Malham Tarn DRT (WMA6) (t-1)  Snaizeholme DRT (WMA6) (t-1)  \
count                   1450.000000                   1450.000000
mean                      13.991815                      9.664587
```

```
std                          14.048930                    12.087704
min                           0.000000                     0.000130
25%                           2.855575                     1.571982
50%                          10.064458                     5.628753
75%                          21.354523                    12.786220
max                         100.117414                    86.247750

       Arkengarthdale DRT (WMA7) (t-1)  East Cowton DRT (WMA7) (t-1)  \
count                      1450.000000                   1450.000000
mean                          5.774654                      5.283198
std                           6.762191                      5.555810
min                           0.000000                      0.000000
25%                           1.227186                      1.504081
50%                           3.792445                      3.664996
75%                           8.049764                      7.160675
max                          77.825716                     45.744930

       Malham Tarn DRT (WMA7) (t-1)  Snaizeholme DRT (WMA7) (t-1)
count                   1450.000000                   1450.000000
mean                      13.993078                      9.673086
std                       13.411499                     11.615246
min                        0.000000                      0.000424
25%                        3.244706                      1.728840
50%                       10.593781                      5.863176
75%                       21.058371                     12.848060
max                       95.347821                     80.502521

[8 rows x 41 columns]
```

[107]:
```
# Exporting datasets with lags and moving averages
lagged_df.to_excel('River-Data-Lagged.xlsx')
moving_avg_df.to_excel('River-Data-MA.xlsx')
lagged_ma_df.to_excel('River-Data-MA-Lagged.xlsx')
weighted_ma_df.to_excel('River-Data-WMA.xlsx')
lagged_wma_df.to_excel('River-Data-WMA-Lagged.xlsx')
```