This notebook features EDA on the AIM 1994 Diabetes Data
data-[01-70]: data sets covering several weeks' to months' worth of outpatient care on 70 patients.

Importing libraries

In [1]:
```
import pandas as pd
import numpy as np
from pathlib import Path
import tarfile
import matplotlib.pyplot as plt
import warnings
import scipy
import seaborn as sns
%matplotlib inline
```

The source is a .zip file, we will extract and store all respective files in our working directory. Since the data belongs to 70 difference outpatients, it is not wise to create a single unified dataframe for exploration. We can load each file using pd.to_csv and iterate a For loop and store the separate tables in a list.

In [ ]:
```
!tar -xvf "C:\Users\FeedSpot.DESKTOP-IQ6759U\Downloads\diabetesdataset\diabetes-data.tar.Z"
```

In [2]:
```
filepath = "Diabetes-Data"
```

In [3]:
```
df_list = []
for i in range(1, 71):
    file_name = f"Diabetes-Data/data-{i:02d}"
    df = pd.read_csv(
        file_name,
        delimiter="\t",            # tab-separated
        header=None,
        names=["Date", "Time", "Code", "Value"]
    )
    df['id']=f"{i:02d}"
    df_list.append(df)
```

Great, now we can access any patient's data by indexing the list

In [ ]:
```
df_list[0].head(6) #data of patient no.1, confirmed by the id column
```

In [ ]:
```
#merged_df = pd.concat(df_list, axis=0, ignore_index=True)
#creating a merged dataframe in case i need it later
```

For now, we will only explore blood glucose trends of patient id=1. If done successfully, it can be re-iterated for other patients

In [4]:
```
df1 = df_list[0]
```

In [5]:
```
len(df1['Value'].isnull())
```

Out[5]: 943

Our aim is to check time series of Blood Glucose values for the patient. So we will only use those rows whose 'Code' feature pertains to glucose measurements

48 = Unspecified blood glucose measurement
57 = Unspecified blood glucose measurement
58 = Pre-breakfast blood glucose measurement
59 = Post-breakfast blood glucose measurement
60 = Pre-lunch blood glucose measurement
61 = Post-lunch blood glucose measurement
62 = Pre-supper blood glucose measurement
63 = Post-supper blood glucose measurement
64 = Pre-snack blood glucose measurement
65 = Hypoglycemic symptoms
Lets create a copy

In [ ]:
```
df1.loc[df1['Code'].isin([65, 66, 67, 68, 69, 70, 71, 72]), 'Value'].head()
```

In [51]:
```
df_bloodglucose = df1[(df1['Code']>47) & (df1['Code']<65)].copy()
```

In [52]:
```
df_bloodglucose.shape
```

Out[52]: (369, 5)

In [53]:
```
df_bloodglucose['datetime'] = pd.to_datetime(df_bloodglucose['Date'] + ' '
+ df_bloodglucose['Time'])
```

In [9]:
```
df_bloodglucose.head()
```

Out[9]:

	Date	Time	Code	Value	id	datetime
0	04-21-1991	9:09	58	100	01	1991-04-21 09:09:00
3	04-21-1991	17:08	62	119	01	1991-04-21 17:08:00
5	04-21-1991	22:51	48	123	01	1991-04-21 22:51:00
6	04-22-1991	7:35	58	216	01	1991-04-22 07:35:00
10	04-22-1991	16:56	62	211	01	1991-04-22 16:56:00

In [54]:
```
df_bloodglucose = df_bloodglucose.drop(['Date', 'Time'], axis=1)
df_bloodglucose.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 369 entries, 0 to 940
Data columns (total 4 columns):
 #   Column    Non-Null Count  Dtype         
---  ------    --------------  -----         
 0   Code      369 non-null    int64         
 1   Value     369 non-null    int64         
 2   id        369 non-null    object        
 3   datetime  369 non-null    datetime64[ns]
dtypes: datetime64[ns](1), int64(2), object(1)
memory usage: 14.4+ KB
```

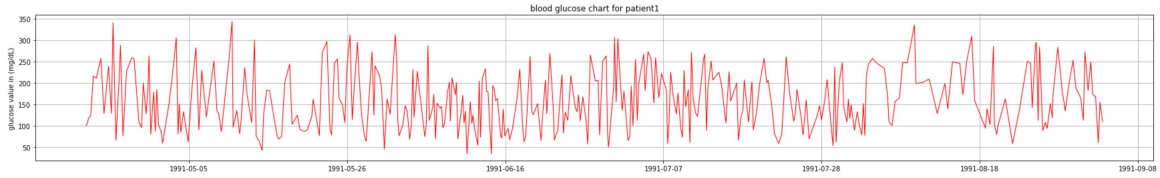
In [55]:
```
df_bloodglucose['datetime'].max() - df_bloodglucose['datetime'].min()
```

Out[55]:
```
Timedelta('134 days 22:11:00')
```

In [56]:
```
df_bloodglucose = df_bloodglucose.sort_values('datetime')
```

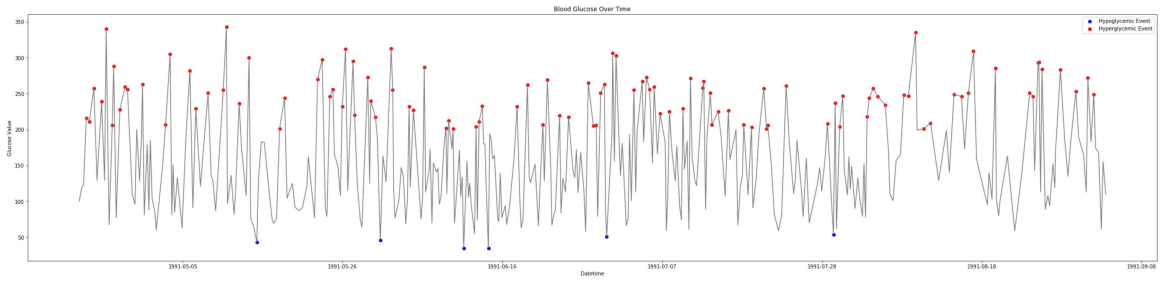
In [57]:
```
import matplotlib.pyplot as plt
import warnings
with warnings.catch_warnings():
    warnings.simplefilter("ignore")
```

In [14]:
```
plt.figure(figsize=(30,4))
plt.plot(df_bloodglucose['datetime'].values,df_bloodglucose['Value'].value
s, color='red', linewidth=1)
plt.title('blood glucose chart for patient1')
plt.xlabel(" ")
plt.ylabel("glucose value in (mg/dL)")
plt.grid(True)
plt.savefig("blood_glucose_trend.png", dpi=300, bbox_inches='tight')
plt.show()
```



Although there exists a specific code '65' for hypoglycemic symptoms in the data, the value for that code drops to zero and hence thats not much useful to us. Instead we will arbitrarily set a cutoff of blood glucose (<55 mg/dl) as an alert point for an hypoglycemic symptom event. Not only that, we will also set 200 as the trigger for an hyperglycemic event

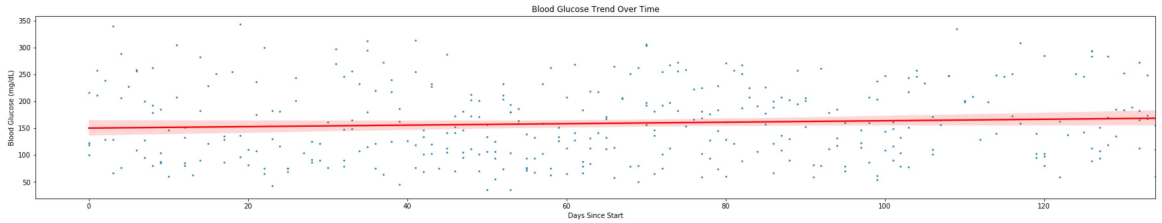
In [63]:
```
plt.figure(figsize=(40,9))
plt.plot(df_bloodglucose['datetime'].values, df_bloodglucose['Value'].value
s, color='grey', linewidth=1.5)


hypo = df_bloodglucose[df_bloodglucose['Value'] < 55]
hyper = df_bloodglucose[df_bloodglucose['Value'] > 200]
plt.scatter(hypo['datetime'], hypo['Value'], color='blue', label='Hypoglyce
mic Event')
plt.scatter(hyper['datetime'], hyper['Value'], color='red', label='Hypergly
cemic Event')

plt.legend()
plt.title('Blood Glucose Over Time')
plt.xlabel('Datetime')
plt.ylabel('Glucose Value')
plt.show()
```



In [16]:
```
df_bloodglucose = df_bloodglucose.set_index('datetime')
df_bloodglucose = df_bloodglucose.sort_index()
```

In [17]:
```
x = (df_bloodglucose.index - df_bloodglucose.index[0]).days.values
y = df_bloodglucose['Value'].values
```

In [18]:
```
plt.figure(figsize=(30,5))
sns.regplot(x=x, y=y, scatter_kws={'s':3, 'alpha':0.9}, line_kws={'colo
r':'red'})
plt.title("Blood Glucose Trend Over Time")
plt.xlabel("Days Since Start")
plt.ylabel("Blood Glucose (mg/dL)")
plt.show()
```

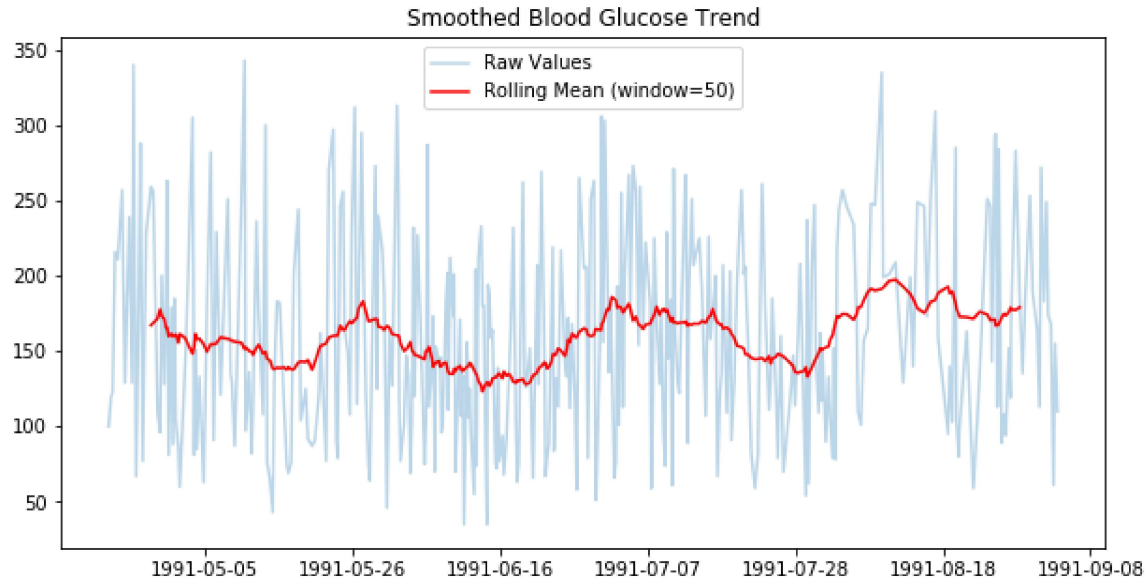


In [19]:
```
slope, intercept = np.polyfit(x, y, 1)
mean = df_bloodglucose['Value'].mean()
print("Slope:", slope )
print("Mean BG for Patient 1:", mean)
if slope > 0.2:
    print ("glucose levels rising, Patient needs attention")
else :
    print ("glucose levels are stable-")
if mean > 150:
    print ("but high, poorly controlled diabetes")
```

```
Slope: 0.13648827164496205
Mean BG for Patient 1: 159.0460704607046
glucose levels are stable-
but high, poorly controlled diabetes
```

Since the average BG doesnt change a lot. We can see the best fit line is mostly flat and hence not much useful.
Let's observe finer trends using a smoother curve

In [20]:
```
df_bloodglucose['rolling_mean'] = df_bloodglucose['Value'].rolling(window=30, center=True).mean()

plt.figure(figsize=(10,5))
plt.plot(df_bloodglucose['Value'], alpha=0.3, label='Raw Values')
plt.plot(df_bloodglucose['rolling_mean'], color='red', label='Rolling Mean (window=50)')
plt.title("Smoothed Blood Glucose Trend")
plt.legend()
plt.show()
```



One can clearly notice the mean fluctuates between 150 and 200, very high values for BG concentration

We have explored Patient 01's data and discovered some insights. Now we need to do it for all 70. This we can do by writing a function containing all the above commands and reiterating it through a loop from 1 till 70

In [100]:
```
from sklearn.linear_model import LinearRegression

def Patient_BG_report(df):
    """
    Analyze a single patient's glucose data.
      - Filtering blood glucose measurements
      - Combining date/time into datetime
      - Linear regression on glucose trend
      - Summary interpretation prints
    """

    # filtering bg measuremnts codes
    df_bg = df[(df['Code'] > 47) & (df['Code'] < 65)].copy()

    if df_bg.empty:
        print("No glucose records found for this patient.")
        return None

   #handling string values
    df_bg['Value'] = df_bg['Value'].replace({'0Hi': '400', '0Lo': '40'})
    df_bg['Value'] = pd.to_numeric(df_bg['Value'], errors='coerce')
    df_bg = df_bg.dropna(subset=['Value'])
    df_bg['Value'] = df_bg['Value'].astype('int64')


    df_bg['Date'] = df_bg['Date'].astype(str).str.strip()
    df_bg['Time'] = df_bg['Time'].astype(str).str.strip()

# Combine and standardize format
    df_bg['datetime_str'] = df_bg['Date'] + ' ' + df_bg['Time']
    df_bg['datetime'] = pd.to_datetime(df_bg['datetime_str'], format='%m-%d
-%Y %H:%M', errors='coerce')


    df_bg = df_bg.dropna(subset=['datetime']).sort_values('datetime')

    # setting days to zero
    df_bg['days_since_start'] = (df_bg['datetime'] - df_bg['datetime'].min
()).dt.days
    x = df_bg['days_since_start'].values.reshape(-1, 1)
    y = df_bg['Value'].values

    # Caluculating slope
    if len(df_bg) > 1:
        model = LinearRegression().fit(x, y)
        slope = model.coef_[0]
    else:
        slope = np.nan

    # mean and standard deviation
    mean_glucose = y.mean()
    std_glucose = y.std()
    pid = df['id'].iloc[0] if 'id' in df.columns else "Unknown"

    # Print Results
    print(f"\n--- Patient {pid} Glucose Report ---")
    print(f"Records: {len(df_bg)}")
    print(f"Mean BG: {mean_glucose:.2f} mg/dL")
    print(f"Slope: {slope:.4f} mg/dL/day")
    print(f"Std Dev: {std_glucose:.2f}")
```

```
    if slope > 0.2:
        print("→ Glucose levels rising: patient may need attention.")


    if mean_glucose > 200:
        print("EXTREMELY HIGH Glucose. Need intervention")
    elif 200 > mean_glucose > 150:
        print("→ High mean glucose — poorly controlled diabetes.")
    elif mean_glucose < 70:
        print("→ Mean glucose too low — risk of hypoglycemia.")
    else:
        print("→ Mean glucose within acceptable range.")

    return { "id": pid,
             "Records":len(df_bg),
             "Mean BG": round(mean_glucose, 2),
             "Slope": round(slope, 4),
             "Std Dev": round(std_glucose, 2) }
```

In [101]:
```
#Let's try a sample
df2 =  df_list[1]
```

In [23]:
```
df2.dtypes
```

Out[23]:
```
Date     object
Time     object
Code      int64
Value    object
id       object
dtype: object
```

When we ran Patient_BG_report(df2) initially, we got an error and discovered that the 'Value' column was not numeric and contained '0Hi' and '0Lo' string values. Our function was corrected and rewritten later to incorporate the non-numeric values across data. However below you will still find the 'manual' fix done to replace values using a dictionary

In [25]:
```
(df2['Value'] == '0Hi').sum()
```

Out[25]:
```
6
```

In [ ]:
```
df2['Value'] = df2['Value'].replace({'0Hi': '400', '0Lo': '40'})
```

In [102]:
```
Patient_BG_report(df2)
```

```
--- Patient 02 Glucose Report ---
Records: 380
Mean BG: 197.01 mg/dL
Slope: -0.1715 mg/dL/day
Std Dev: 79.24
→ High mean glucose — poorly controlled diabetes.
```

Out[102]:
```
{'id': '02',
 'Records': 380,
 'Mean BG': 197.01,
 'Slope': -0.1715,
 'Std Dev': 79.24}
```

In [64]:
```
#Lets try another sample
df6 =  df_list[5]
```

In [111]:
```
Patient_BG_report(df6)
```

```
--- Patient 06 Glucose Report ---
Records: 65
Mean BG: 200.74 mg/dL
Slope: 0.0923 mg/dL/day
Std Dev: 60.51
EXTREMELY HIGH Glucose. Need intervention
```

Out[111]:
```
{'id': '06',
 'Records': 65,
 'Mean BG': 200.74,
 'Slope': 0.0923,
 'Std Dev': 60.51}
```

In [ ]:
```
#It works!
```

Now let's create a loop to run our function over and over to generate the blood glucose analysis and trend reports for all 70 patients

In [103]:
```
results = []

for i, df in enumerate(df_list, start=1):
    print(f"\nProcessing Patient {i:02d}...")
    result = Patient_BG_report(df)
    if result is not None:
            results.append(result)
```

```
Processing Patient 01...

--- Patient 01 Glucose Report ---
Records: 369
Mean BG: 159.05 mg/dL
Slope: 0.1365 mg/dL/day
Std Dev: 69.26
→ High mean glucose — poorly controlled diabetes.

Processing Patient 02...

--- Patient 02 Glucose Report ---
Records: 380
Mean BG: 197.01 mg/dL
Slope: -0.1715 mg/dL/day
Std Dev: 79.24
→ High mean glucose — poorly controlled diabetes.

Processing Patient 03...

--- Patient 03 Glucose Report ---
Records: 149
Mean BG: 135.37 mg/dL
Slope: 0.4683 mg/dL/day
Std Dev: 64.38
→ Glucose levels rising: patient may need attention.
→ Mean glucose within acceptable range.

Processing Patient 04...

--- Patient 04 Glucose Report ---
Records: 147
Mean BG: 162.56 mg/dL
Slope: 0.7356 mg/dL/day
Std Dev: 78.46
→ Glucose levels rising: patient may need attention.
→ High mean glucose — poorly controlled diabetes.

Processing Patient 05...

--- Patient 05 Glucose Report ---
Records: 150
Mean BG: 167.77 mg/dL
Slope: 0.2720 mg/dL/day
Std Dev: 78.47
→ Glucose levels rising: patient may need attention.
→ High mean glucose — poorly controlled diabetes.

Processing Patient 06...

--- Patient 06 Glucose Report ---
Records: 65
Mean BG: 200.74 mg/dL
Slope: 0.0923 mg/dL/day
Std Dev: 60.51
EXTREMELY HIGH Glucose. Need intervention

Processing Patient 07...

--- Patient 07 Glucose Report ---
Records: 113
```

```
Mean BG: 174.51 mg/dL
Slope: -0.0398 mg/dL/day
Std Dev: 63.07
→ High mean glucose — poorly controlled diabetes.

Processing Patient 08...

--- Patient 08 Glucose Report ---
Records: 139
Mean BG: 182.56 mg/dL
Slope: 0.4190 mg/dL/day
Std Dev: 87.74
→ Glucose levels rising: patient may need attention.
→ High mean glucose — poorly controlled diabetes.

Processing Patient 09...

--- Patient 09 Glucose Report ---
Records: 99
Mean BG: 167.18 mg/dL
Slope: 0.2501 mg/dL/day
Std Dev: 94.23
→ Glucose levels rising: patient may need attention.
→ High mean glucose — poorly controlled diabetes.

Processing Patient 10...

--- Patient 10 Glucose Report ---
Records: 107
Mean BG: 200.54 mg/dL
Slope: 0.3255 mg/dL/day
Std Dev: 66.39
→ Glucose levels rising: patient may need attention.
EXTREMELY HIGH Glucose. Need intervention

Processing Patient 11...

--- Patient 11 Glucose Report ---
Records: 68
Mean BG: 147.71 mg/dL
Slope: -0.3812 mg/dL/day
Std Dev: 95.18
→ Mean glucose within acceptable range.

Processing Patient 12...

--- Patient 12 Glucose Report ---
Records: 90
Mean BG: 162.90 mg/dL
Slope: 0.5323 mg/dL/day
Std Dev: 104.57
→ Glucose levels rising: patient may need attention.
→ High mean glucose — poorly controlled diabetes.

Processing Patient 13...

--- Patient 13 Glucose Report ---
Records: 97
Mean BG: 160.62 mg/dL
Slope: 1.1771 mg/dL/day
Std Dev: 103.56
```

```
→ Glucose levels rising: patient may need attention.
→ High mean glucose — poorly controlled diabetes.

Processing Patient 14...

--- Patient 14 Glucose Report ---
Records: 58
Mean BG: 159.79 mg/dL
Slope: 0.4275 mg/dL/day
Std Dev: 95.06
→ Glucose levels rising: patient may need attention.
→ High mean glucose — poorly controlled diabetes.

Processing Patient 15...

--- Patient 15 Glucose Report ---
Records: 93
Mean BG: 158.56 mg/dL
Slope: -0.4010 mg/dL/day
Std Dev: 96.68
→ High mean glucose — poorly controlled diabetes.

Processing Patient 16...

--- Patient 16 Glucose Report ---
Records: 93
Mean BG: 179.38 mg/dL
Slope: 1.3967 mg/dL/day
Std Dev: 101.27
→ Glucose levels rising: patient may need attention.
→ High mean glucose — poorly controlled diabetes.

Processing Patient 17...

--- Patient 17 Glucose Report ---
Records: 122
Mean BG: 141.98 mg/dL
Slope: -0.2682 mg/dL/day
Std Dev: 73.11
→ Mean glucose within acceptable range.

Processing Patient 18...

--- Patient 18 Glucose Report ---
Records: 136
Mean BG: 141.54 mg/dL
Slope: -1.0563 mg/dL/day
Std Dev: 74.53
→ Mean glucose within acceptable range.

Processing Patient 19...

--- Patient 19 Glucose Report ---
Records: 129
Mean BG: 159.81 mg/dL
Slope: 1.7309 mg/dL/day
Std Dev: 88.88
→ Glucose levels rising: patient may need attention.
→ High mean glucose — poorly controlled diabetes.

Processing Patient 20...
```

```
--- Patient 20 Glucose Report ---
Records: 451
Mean BG: 173.10 mg/dL
Slope: -0.1892 mg/dL/day
Std Dev: 93.37
→ High mean glucose — poorly controlled diabetes.

Processing Patient 21...

--- Patient 21 Glucose Report ---
Records: 207
Mean BG: 144.00 mg/dL
Slope: 0.1640 mg/dL/day
Std Dev: 47.11
→ Mean glucose within acceptable range.

Processing Patient 22...

--- Patient 22 Glucose Report ---
Records: 104
Mean BG: 181.61 mg/dL
Slope: -0.6457 mg/dL/day
Std Dev: 99.27
→ High mean glucose — poorly controlled diabetes.

Processing Patient 23...

--- Patient 23 Glucose Report ---
Records: 103
Mean BG: 172.17 mg/dL
Slope: -0.5284 mg/dL/day
Std Dev: 101.31
→ High mean glucose — poorly controlled diabetes.

Processing Patient 24...

--- Patient 24 Glucose Report ---
Records: 117
Mean BG: 164.97 mg/dL
Slope: 0.2761 mg/dL/day
Std Dev: 96.56
→ Glucose levels rising: patient may need attention.
→ High mean glucose — poorly controlled diabetes.

Processing Patient 25...

--- Patient 25 Glucose Report ---
Records: 92
Mean BG: 123.79 mg/dL
Slope: 1.7057 mg/dL/day
Std Dev: 62.90
→ Glucose levels rising: patient may need attention.
→ Mean glucose within acceptable range.

Processing Patient 26...

--- Patient 26 Glucose Report ---
Records: 305
Mean BG: 137.55 mg/dL
Slope: -0.2605 mg/dL/day
```

```
Std Dev: 59.98
→ Mean glucose within acceptable range.

Processing Patient 27...

--- Patient 27 Glucose Report ---
Records: 457
Mean BG: 127.03 mg/dL
Slope: 0.0769 mg/dL/day
Std Dev: 44.86
→ Mean glucose within acceptable range.

Processing Patient 28...

--- Patient 28 Glucose Report ---
Records: 475
Mean BG: 141.68 mg/dL
Slope: -0.1037 mg/dL/day
Std Dev: 53.79
→ Mean glucose within acceptable range.

Processing Patient 29...

--- Patient 29 Glucose Report ---
Records: 616
Mean BG: 160.95 mg/dL
Slope: 0.1131 mg/dL/day
Std Dev: 44.87
→ High mean glucose — poorly controlled diabetes.

Processing Patient 30...

--- Patient 30 Glucose Report ---
Records: 591
Mean BG: 152.63 mg/dL
Slope: 0.0225 mg/dL/day
Std Dev: 51.11
→ High mean glucose — poorly controlled diabetes.

Processing Patient 31...

--- Patient 31 Glucose Report ---
Records: 335
Mean BG: 146.83 mg/dL
Slope: 0.6310 mg/dL/day
Std Dev: 46.98
→ Glucose levels rising: patient may need attention.
→ Mean glucose within acceptable range.

Processing Patient 32...

--- Patient 32 Glucose Report ---
Records: 62
Mean BG: 158.50 mg/dL
Slope: -3.3402 mg/dL/day
Std Dev: 77.43
→ High mean glucose — poorly controlled diabetes.

Processing Patient 33...

--- Patient 33 Glucose Report ---
```

```
Records: 121
Mean BG: 153.98 mg/dL
Slope: 0.3749 mg/dL/day
Std Dev: 67.75
→ Glucose levels rising: patient may need attention.
→ High mean glucose — poorly controlled diabetes.

Processing Patient 34...

--- Patient 34 Glucose Report ---
Records: 121
Mean BG: 148.50 mg/dL
Slope: -0.0831 mg/dL/day
Std Dev: 60.62
→ Mean glucose within acceptable range.

Processing Patient 35...

--- Patient 35 Glucose Report ---
Records: 122
Mean BG: 140.19 mg/dL
Slope: -0.7189 mg/dL/day
Std Dev: 65.45
→ Mean glucose within acceptable range.

Processing Patient 36...

--- Patient 36 Glucose Report ---
Records: 124
Mean BG: 173.89 mg/dL
Slope: -0.0725 mg/dL/day
Std Dev: 85.52
→ High mean glucose — poorly controlled diabetes.

Processing Patient 37...

--- Patient 37 Glucose Report ---
Records: 131
Mean BG: 178.17 mg/dL
Slope: -0.5143 mg/dL/day
Std Dev: 76.72
→ High mean glucose — poorly controlled diabetes.

Processing Patient 38...

--- Patient 38 Glucose Report ---
Records: 115
Mean BG: 162.15 mg/dL
Slope: 2.4884 mg/dL/day
Std Dev: 82.11
→ Glucose levels rising: patient may need attention.
→ High mean glucose — poorly controlled diabetes.

Processing Patient 39...

--- Patient 39 Glucose Report ---
Records: 119
Mean BG: 149.61 mg/dL
Slope: -1.6815 mg/dL/day
Std Dev: 75.42
→ Mean glucose within acceptable range.
```

```
Processing Patient 40...

--- Patient 40 Glucose Report ---
Records: 31
Mean BG: 136.97 mg/dL
Slope: 0.6305 mg/dL/day
Std Dev: 93.16
→ Glucose levels rising: patient may need attention.
→ Mean glucose within acceptable range.

Processing Patient 41...

--- Patient 41 Glucose Report ---
Records: 243
Mean BG: 167.09 mg/dL
Slope: -0.4226 mg/dL/day
Std Dev: 76.50
→ High mean glucose — poorly controlled diabetes.

Processing Patient 42...

--- Patient 42 Glucose Report ---
Records: 254
Mean BG: 146.75 mg/dL
Slope: 0.0197 mg/dL/day
Std Dev: 83.54
→ Mean glucose within acceptable range.

Processing Patient 43...

--- Patient 43 Glucose Report ---
Records: 121
Mean BG: 154.04 mg/dL
Slope: 2.0581 mg/dL/day
Std Dev: 87.11
→ Glucose levels rising: patient may need attention.
→ High mean glucose — poorly controlled diabetes.

Processing Patient 44...

--- Patient 44 Glucose Report ---
Records: 104
Mean BG: 165.67 mg/dL
Slope: 1.0660 mg/dL/day
Std Dev: 88.55
→ Glucose levels rising: patient may need attention.
→ High mean glucose — poorly controlled diabetes.

Processing Patient 45...

--- Patient 45 Glucose Report ---
Records: 101
Mean BG: 167.34 mg/dL
Slope: -0.7631 mg/dL/day
Std Dev: 88.71
→ High mean glucose — poorly controlled diabetes.

Processing Patient 46...

--- Patient 46 Glucose Report ---
```

```
Records: 95
Mean BG: 164.13 mg/dL
Slope: -0.1888 mg/dL/day
Std Dev: 81.47
→ High mean glucose — poorly controlled diabetes.

Processing Patient 47...

--- Patient 47 Glucose Report ---
Records: 101
Mean BG: 132.46 mg/dL
Slope: -0.9288 mg/dL/day
Std Dev: 83.35
→ Mean glucose within acceptable range.

Processing Patient 48...

--- Patient 48 Glucose Report ---
Records: 104
Mean BG: 160.83 mg/dL
Slope: -0.3834 mg/dL/day
Std Dev: 80.82
→ High mean glucose — poorly controlled diabetes.

Processing Patient 49...

--- Patient 49 Glucose Report ---
Records: 211
Mean BG: 94.00 mg/dL
Slope: -0.1110 mg/dL/day
Std Dev: 38.93
→ Mean glucose within acceptable range.

Processing Patient 50...

--- Patient 50 Glucose Report ---
Records: 277
Mean BG: 87.09 mg/dL
Slope: -0.6098 mg/dL/day
Std Dev: 33.54
→ Mean glucose within acceptable range.

Processing Patient 51...

--- Patient 51 Glucose Report ---
Records: 157
Mean BG: 112.99 mg/dL
Slope: -0.0541 mg/dL/day
Std Dev: 59.89
→ Mean glucose within acceptable range.

Processing Patient 52...

--- Patient 52 Glucose Report ---
Records: 175
Mean BG: 131.28 mg/dL
Slope: 1.8106 mg/dL/day
Std Dev: 88.84
→ Glucose levels rising: patient may need attention.
→ Mean glucose within acceptable range.
```

```
Processing Patient 53...

--- Patient 53 Glucose Report ---
Records: 123
Mean BG: 166.46 mg/dL
Slope: -0.5345 mg/dL/day
Std Dev: 73.31
→ High mean glucose — poorly controlled diabetes.

Processing Patient 54...

--- Patient 54 Glucose Report ---
Records: 362
Mean BG: 139.99 mg/dL
Slope: 0.3593 mg/dL/day
Std Dev: 69.61
→ Glucose levels rising: patient may need attention.
→ Mean glucose within acceptable range.

Processing Patient 55...

--- Patient 55 Glucose Report ---
Records: 606
Mean BG: 157.87 mg/dL
Slope: -0.3245 mg/dL/day
Std Dev: 65.48
→ High mean glucose — poorly controlled diabetes.

Processing Patient 56...

--- Patient 56 Glucose Report ---
Records: 499
Mean BG: 177.03 mg/dL
Slope: -0.0917 mg/dL/day
Std Dev: 94.93
→ High mean glucose — poorly controlled diabetes.

Processing Patient 57...

--- Patient 57 Glucose Report ---
Records: 103
Mean BG: 215.40 mg/dL
Slope: 0.1055 mg/dL/day
Std Dev: 113.17
EXTREMELY HIGH Glucose. Need intervention

Processing Patient 58...

--- Patient 58 Glucose Report ---
Records: 116
Mean BG: 218.65 mg/dL
Slope: -0.5171 mg/dL/day
Std Dev: 103.02
EXTREMELY HIGH Glucose. Need intervention

Processing Patient 59...

--- Patient 59 Glucose Report ---
Records: 120
Mean BG: 208.66 mg/dL
Slope: -0.9990 mg/dL/day
```

```
Std Dev: 98.45
EXTREMELY HIGH Glucose. Need intervention

Processing Patient 60...

--- Patient 60 Glucose Report ---
Records: 112
Mean BG: 133.05 mg/dL
Slope: 1.5925 mg/dL/day
Std Dev: 77.14
→ Glucose levels rising: patient may need attention.
→ Mean glucose within acceptable range.

Processing Patient 61...

--- Patient 61 Glucose Report ---
Records: 120
Mean BG: 201.47 mg/dL
Slope: 0.8051 mg/dL/day
Std Dev: 112.21
→ Glucose levels rising: patient may need attention.
EXTREMELY HIGH Glucose. Need intervention

Processing Patient 62...

--- Patient 62 Glucose Report ---
Records: 132
Mean BG: 201.55 mg/dL
Slope: -0.8871 mg/dL/day
Std Dev: 109.04
EXTREMELY HIGH Glucose. Need intervention

Processing Patient 63...

--- Patient 63 Glucose Report ---
Records: 138
Mean BG: 197.58 mg/dL
Slope: -0.2247 mg/dL/day
Std Dev: 106.87
→ High mean glucose — poorly controlled diabetes.

Processing Patient 64...

--- Patient 64 Glucose Report ---
Records: 82
Mean BG: 121.67 mg/dL
Slope: -1.4230 mg/dL/day
Std Dev: 69.62
→ Mean glucose within acceptable range.

Processing Patient 65...

--- Patient 65 Glucose Report ---
Records: 546
Mean BG: 208.60 mg/dL
Slope: 0.1317 mg/dL/day
Std Dev: 77.23
EXTREMELY HIGH Glucose. Need intervention

Processing Patient 66...
```

```
--- Patient 66 Glucose Report ---
Records: 87
Mean BG: 207.39 mg/dL
Slope: -0.4085 mg/dL/day
Std Dev: 73.68
EXTREMELY HIGH Glucose. Need intervention

Processing Patient 67...

--- Patient 67 Glucose Report ---
Records: 465
Mean BG: 209.18 mg/dL
Slope: 0.2083 mg/dL/day
Std Dev: 92.29
→ Glucose levels rising: patient may need attention.
EXTREMELY HIGH Glucose. Need intervention

Processing Patient 68...

--- Patient 68 Glucose Report ---
Records: 527
Mean BG: 121.65 mg/dL
Slope: -0.0173 mg/dL/day
Std Dev: 33.11
→ Mean glucose within acceptable range.

Processing Patient 69...

--- Patient 69 Glucose Report ---
Records: 51
Mean BG: 205.25 mg/dL
Slope: -1.1549 mg/dL/day
Std Dev: 97.45
EXTREMELY HIGH Glucose. Need intervention

Processing Patient 70...

--- Patient 70 Glucose Report ---
Records: 109
Mean BG: 219.03 mg/dL
Slope: 0.5738 mg/dL/day
Std Dev: 106.67
→ Glucose levels rising: patient may need attention.
EXTREMELY HIGH Glucose. Need intervention
```

In [105]:
```
#Let's summarize the patient's bg data in a single data frame
summary_df = pd.DataFrame(results)
```

In [106]:
```
summary_df.head(6)
```

Out[106]:

	id	Records	Mean BG	Slope	Std Dev
0	01	369	159.05	0.1365	69.26
1	02	380	197.01	-0.1715	79.24
2	03	149	135.37	0.4683	64.38
3	04	147	162.56	0.7356	78.46
4	05	150	167.77	0.2720	78.47
5	06	65	200.74	0.0923	60.51

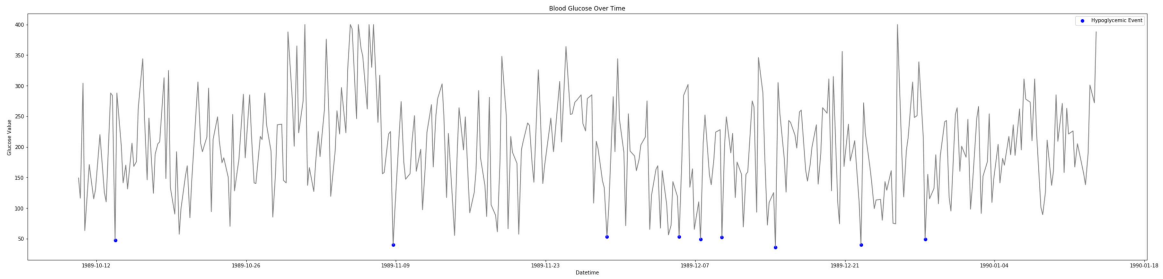
So far, we have managed to analyse and club together all the patients' bg data togther numerically. The dataframe summary_df does not help us visualize how the bg line moves though.

Let's write another function to generate a patient's blood glucose plot. we will also annotate hypoglycemic symptoms.

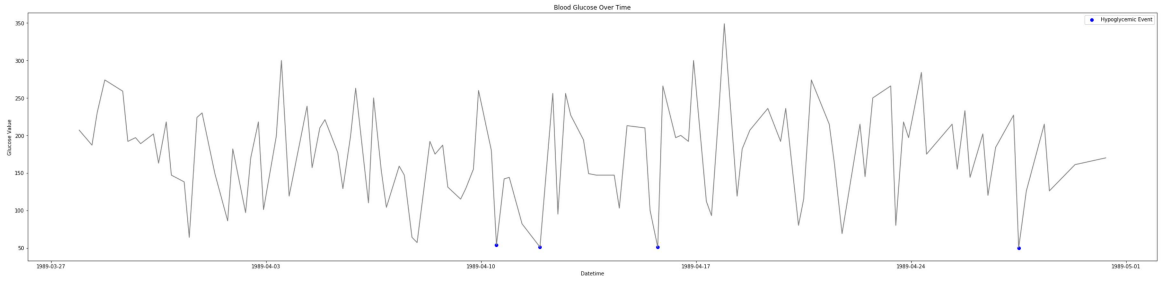
In [114]:
```
def Blood_Glucose_Chart(df):

    df_bg = df[(df['Code'] > 47) & (df['Code'] < 65)].copy()

    if df_bg.empty:
        print("No glucose records found for this patient.")
        return None

   #handling string values
    df_bg['Value'] = df_bg['Value'].replace({'0Hi': '400', '0Lo': '40'})
    df_bg['Value'] = pd.to_numeric(df_bg['Value'], errors='coerce')
    df_bg = df_bg.dropna(subset=['Value'])
    df_bg['Value'] = df_bg['Value'].astype('int64')


    df_bg['Date'] = df_bg['Date'].astype(str).str.strip()
    df_bg['Time'] = df_bg['Time'].astype(str).str.strip()

# Combine and standardize format
    df_bg['datetime_str'] = df_bg['Date'] + ' ' + df_bg['Time']

# Try to parse, coercing invalid ones to NaT
    df_bg['datetime'] = pd.to_datetime(df_bg['datetime_str'], format='%m-%d
-%Y %H:%M', errors='coerce')


    df_bg = df_bg.dropna(subset=['datetime']).sort_values('datetime')

    # setting days to zero
    df_bg['days_since_start'] = (df_bg['datetime'] - df_bg['datetime'].min
()).dt.days
    x = df_bg['days_since_start'].values.reshape(-1, 1)
    y = df_bg['Value'].values

    pid = df['id'].iloc[0] if 'id' in df.columns else "Unknown"


    plt.figure(figsize=(40,9))
    plt.plot(df_bg['datetime'].values, df_bg['Value'].values, color='grey',
linewidth=1.5)


    hypo = df_bg[df_bg['Value'] < 55]

    plt.scatter(hypo['datetime'], hypo['Value'], color='blue', label='Hypog
lycemic Event')
```

```
    plt.legend()
    plt.title('Blood Glucose Over Time')
    plt.xlabel('Datetime')
    plt.ylabel('Glucose Value')
    plt.show()
```

In [115]:
```
#Let's try a few samples
Blood_Glucose_Chart(df2)
```



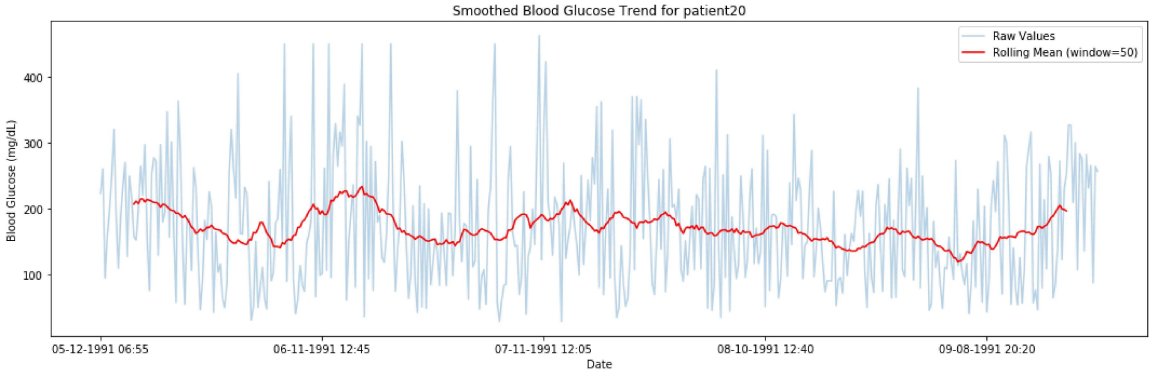
In [118]:
```
df7 = df_list[6]
Blood_Glucose_Chart(df7)
```



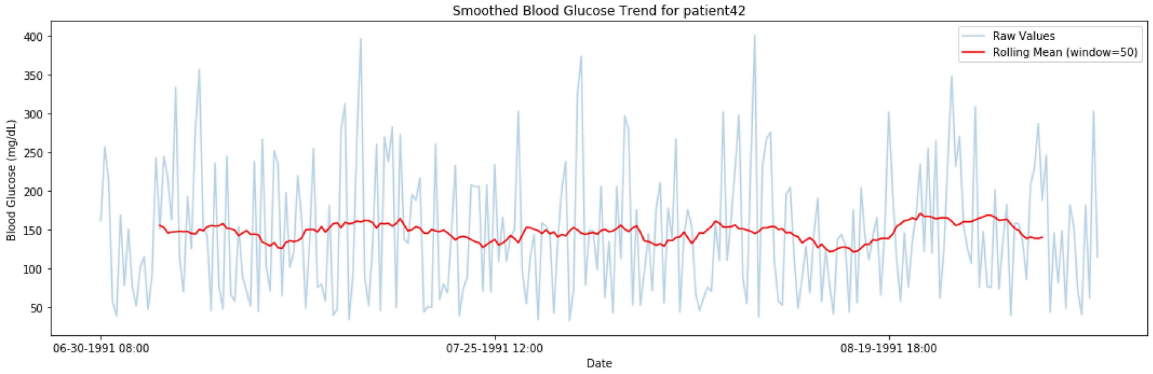
Now as a final exercise let's write a function to generate a Smoothed Blood Glucose Trend

In [169]:
```
def BloodGlucoseTrendOverTime(df):

    df_bg = df[(df['Code'] > 47) & (df['Code'] < 65)].copy()

    if df_bg.empty:
        print("No glucose records found for this patient.")
        return None

   #handling string values
    df_bg['Value'] = df_bg['Value'].replace({'0Hi': '400', '0Lo': '40'})
    df_bg['Value'] = pd.to_numeric(df_bg['Value'], errors='coerce')
    df_bg = df_bg.dropna(subset=['Value'])
    df_bg['Value'] = df_bg['Value'].astype('int64')


    df_bg['Date'] = df_bg['Date'].astype(str).str.strip()
    df_bg['Time'] = df_bg['Time'].astype(str).str.strip()

# Combine and standardize format
    df_bg['datetime_str'] = df_bg['Date'] + ' ' + df_bg['Time']

# Try to parse, coercing invalid ones to NaT
    df_bg['datetime'] = pd.to_datetime(df_bg['datetime_str'], format='%m-%d
-%Y %H:%M', errors='coerce')


    df_bg = df_bg.dropna(subset=['datetime']).sort_values('datetime')

    # setting days to zero
    #df_bg['days_since_start'] = (df_bg['datetime'] - df_bg['datetime'].min
()).dt.days
    #x = df_bg['days_since_start'].values.reshape(-1, 1)
    #y = df_bg['Value'].values
    #No we are not
    #x = (df_bg.index - df_bg.index[0]).days.values
    #y = df_bg['Value'].values

    pid = df['id'].iloc[0] if 'id' in df.columns else "Unknown"

    df_bg['rolling_mean'] = df_bg['Value'].rolling(window=30, center=True).
mean()

    plt.figure(figsize=(15,5))
    plt.plot(df_bg['datetime_str'], df_bg['Value'], alpha=0.3, label='Raw V
alues')
    plt.plot(df_bg['datetime_str'], df_bg['rolling_mean'], color='red', lab
el='Rolling Mean (window=50)')
    plt.title(f"Smoothed Blood Glucose Trend for patient{pid}")
    plt.legend()
    plt.xticks(df_bg['datetime_str'][::100], rotation=0)   # every 200th la
bel, adjust as needed
    plt.xlabel("Date")
    plt.ylabel("Blood Glucose (mg/dL)")
    plt.tight_layout()
    plt.show()
```

In [170]:
```
#Samples
BloodGlucoseTrendOverTime(df20)
```



In [171]:
```
df42 = df_list[41]
BloodGlucoseTrendOverTime(df42)
```



EDA or Exploratory Data Analysis in an important step in handling any Dataset. Due to the nature of the data given, we have managed here to separate, organize and analyze 70 different 'minisets' with concise numeric as well as visual for better understanding of the data. Cheerios!

In [ ]:

In [ ]: