Palliuds-MR1-Analysis

[32]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
from scipy.stats import ttest_ind, pearsonr
from scipy.stats import ttest_rel
from IPython.display import Image
```

0.0.1 Data Loading

[23]:
```python
data = pd.read_excel('Pallidus-MR1-Data.xlsx', parse_dates=['Date & Time'])
```

[24]:
```python
data.head(2)
```

[24]:
```
           Date & Time  0002CC01000006DC  0002CC010000058B  \
0 2023-01-17 06:00:00              2.04              0.02
1 2023-01-17 06:30:00              6.56              0.02

   0002CC0100000708  0002CC0100000664
0              4.24              0.29
1              0.60              3.87
```

0.0.2 Data preprocessing

[25]:
```python
detected_devices = list(data.columns[1:])

data.set_index("Date & Time", drop=False, inplace=True)
data['Hour'] = data['Date & Time'].dt.hour
data['Day'] = data['Date & Time'].dt.day_name()
data["timeDay"] = data["Hour"].apply(lambda x: "day" if 6 + 1 <= x < 18 else␣
↪"night")
```

[26]:
```python
# day and night periods
day_period = data[(data['Hour'] >= 6) & (data['Hour'] < 18)]
night_period = data[(data['Hour'] >= 18) | (data['Hour'] < 6)]
```

0.1 Activity overtime

This section visualizes the activity data over time, highlighting the day and night periods. This helps in identifying patterns and differences in activity during the day and night, which can be crucial for understanding the behavior and habits of the subjects. By clearly distinguishing these day and night periods, the plot will help in analysing peaks and drops of activities over the time alongwith comparison of subject' activities.

[27]:
```
colors_list = ["#264653", "#2a9d8f", "#e9c46a", "#f4a261", "#e76f51",␣
↪"#457b9d", "#fb6f92"]

fig = go.Figure()
ind=0
for i in detected_devices:
    fig.add_trace(
        go.Scatter(
            x=data["Date & Time"], y=data[i], mode='lines',
            marker={'color': colors_list[ind]}, name=str(i)[-3:]
        )
    )
    ind += 1

start_day_hour, end_day_hour = (6, 18)

if (data["Hour"] == 0).sum() != data.shape[0]:
    ind = data.timeDay.index
    vals = data.timeDay.values
    prev = 0
    times = []
    for i in range(1, data.timeDay.size - 1):
        if vals[i] != vals[prev]:
            if vals[prev] == "night":
                times.append((ind[prev], ind[i - 1]))
            prev = i
    if vals[-1] == "night":
        times.append((ind[prev], ind[-1]))
    for time in times:
        fig.add_vrect(
            x0=time[0],
            x1=time[1],
            fillcolor="black",
            opacity=0.3,
            line_width=0,
        )
fig.update_layout(yaxis_title="Activity", xaxis_title="Time", hovermode="x␣
↪unified",
                      hoverlabel=dict(bgcolor="gray",␣
↪font_color="black",font_size=16, font_family="Rockwell"),
```

```
                                xaxis=dict(rangeslider=dict(visible=True)),
                                height=600
                                )
fig.show()
```

[36]:
```
Image(filename='plot1.png')
```

[36]:

From the plot, it can be seen that the activities seem to be higher during the dark cycle and lower during the light cycle. It clearly indicates that the subjects are more active; under the conditions(in which data is collected), during the night as compared to day time.

0.1.1 Day vs Night activity

The next concern is to find the average activity levels during day and night for each detected device. By grouping the data based on time of day, it highlights the differences in activity patterns between daytime and nighttime. The bar chart representation will be helpful here, providing insights into how activity levels vary with the time of day. This analysis will helps in understanding the behavior patterns and potential circadian rhythms of the subjects.

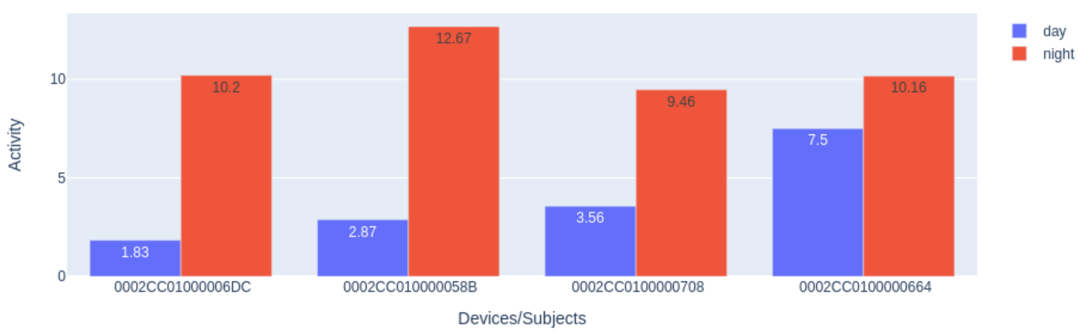
[28]:
```
day_night_activity = data.groupby("timeDay")[detected_devices].mean().
  ↪transpose()
```

[29]:
```
fig = go.Figure()

for i in day_night_activity.columns:
```

```
    fig.add_trace(
        go.Bar(
            x=day_night_activity.index, y=day_night_activity[i], name=i,␣
 ↪text=round(day_night_activity[i],2)
        )
    )
fig.update_layout(xaxis_title="Devices/Subjects", yaxis_title="Activity",␣
 ↪height=400)
fig.show()
```

[34]:
```
Image(filename='plot2.png')
```

[34]:



The chart above compares the average activity levels during the day and night for each detected device. Analysing this we can conclude:

1. **Overall Higher Night Activity:** The activity levels for all devices are significantly higher during the night compared to the day. This suggests that the subjects are more active at night, which could be indicative of nocturnal behavior.
2. **Device 0002CC0100000664:** This device shows the highest activity levels both during the day (7.50) and night (10.16), indicating a consistently active subject.
3. **Device 0002CC010000058B:** This device has a notable increase from day activity (2.87) to night activity (12.67), showing the most significant difference in activity levels between day and night.
4. **Consistent Patterns:** All devices show a similar pattern of increased activity at night, which could be reflective of the natural circadian rhythms of the subjects, suggesting a tendency to be more active during the nighttime.

0.2 Hypothesis Testing

0.2.1 Hypothesis 1:

Subjects exhibit higher activity levels during the night compared to the day.

Hypothesis Formulation: Null Hypothesis (H0): *There is no significant difference in the activity levels of subjects between day and night.*

$ H_0 : *{day}* = {night} $

Alternate Hypothesis (H1): *There is a significant difference in the activity levels of subjects between day and night.*

$ H_1 : *{day}* {night} $

This following test will compare the mean activity levels during the day and night for each device to determine if the difference is statistically significant.

[9]:
```python
day_data = data[data["timeDay"] == "day"][detected_devices]
night_data = data[data["timeDay"] == "night"][detected_devices]

day_means = day_data.mean()
night_means = night_data.mean()
```

We will be using ttest_rel (paired t-test) because the day and night activity measurements for each device are related(or paired). Each device provides two measurements: one for day and one for night, so the data points are not independent of each other.

[10]:
```python
# Perform paired t-test
t_stat, p_value = ttest_rel(day_means, night_means)

# Print the results
print(f"T-statistic: {t_stat}")
print(f"P-value: {p_value}")

# Interpretation
alpha = 0.05
if p_value < alpha:
    print("Reject the null hypothesis: Subjects exhibit higher activity levels during the night compared to the day.")
else:
    print("Fail to reject the null hypothesis: There is no significant difference in activity levels between day and night.")
```

```
T-statistic: -4.2772092626351865
P-value: 0.023469759976103274
Reject the null hypothesis: Subjects exhibit higher activity levels during the
night compared to the day.
```

The t-stat value of -4.28 and a p-value of 0.023 indicate a statistically significant difference in activity levels between day and night. Since the p-value is less than the significance level of 0.05, we reject the null hypothesis. This suggests that subjects exhibit significantly higher activity levels during the night compared to the day.

0.2.2 Hypothesis 2 : Influence of One Subject's Activity on Others

The activity of one subject influences the activity of others.

Hypothesis Formulation: **Null Hypothesis (H0):** *There is no correlation between the activity levels of one subject and the activity levels of other subjects.*

Alternate Hypothesis (H1): *There is a significant correlation between the activity levels of one subject and the activity levels of other subjects.*

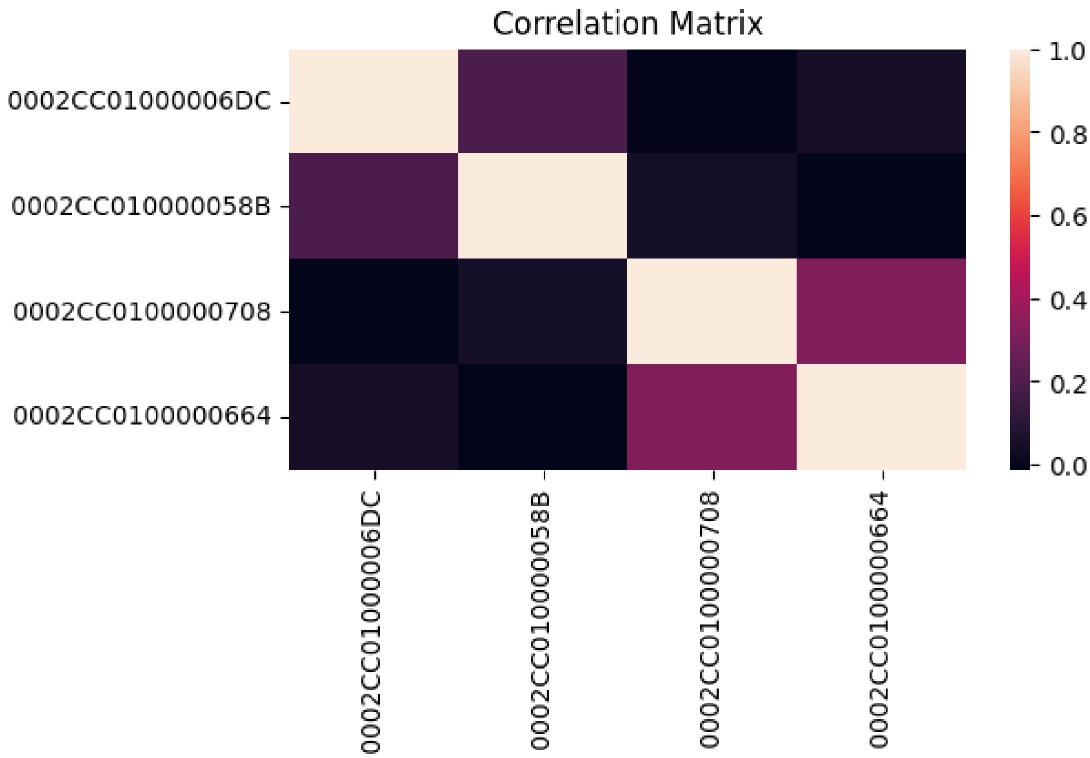
This hypothesis test will examine whether the activity level of one subject is correlated with the activity levels of other subjects. By analyzing the correlation between pairs of subjects' activity data, we can determine if there is a significant relationship suggesting that the activity of one subject influences the activity of others.

[11]:
```
data_test = data.dropna()

# pairwise correlations btw activity levels
correlations = data_test[detected_devices].corr()

# Perform Pearson correlation test
p_values = pd.DataFrame(index=detected_devices, columns=detected_devices)
```

[12]:
```
for i in detected_devices:
    for j in detected_devices:
        if i != j:
            corr, p_value = pearsonr(data_test[i], data_test[j])
            p_values.loc[i, j] = p_value
```

[13]:
```
plt.figure(figsize=(6,3))
sns.heatmap(correlations)
plt.title("Correlation Matrix")
```

[13]:
```
Text(0.5, 1.0, 'Correlation Matrix')
```



[14]:
```
p_values
```

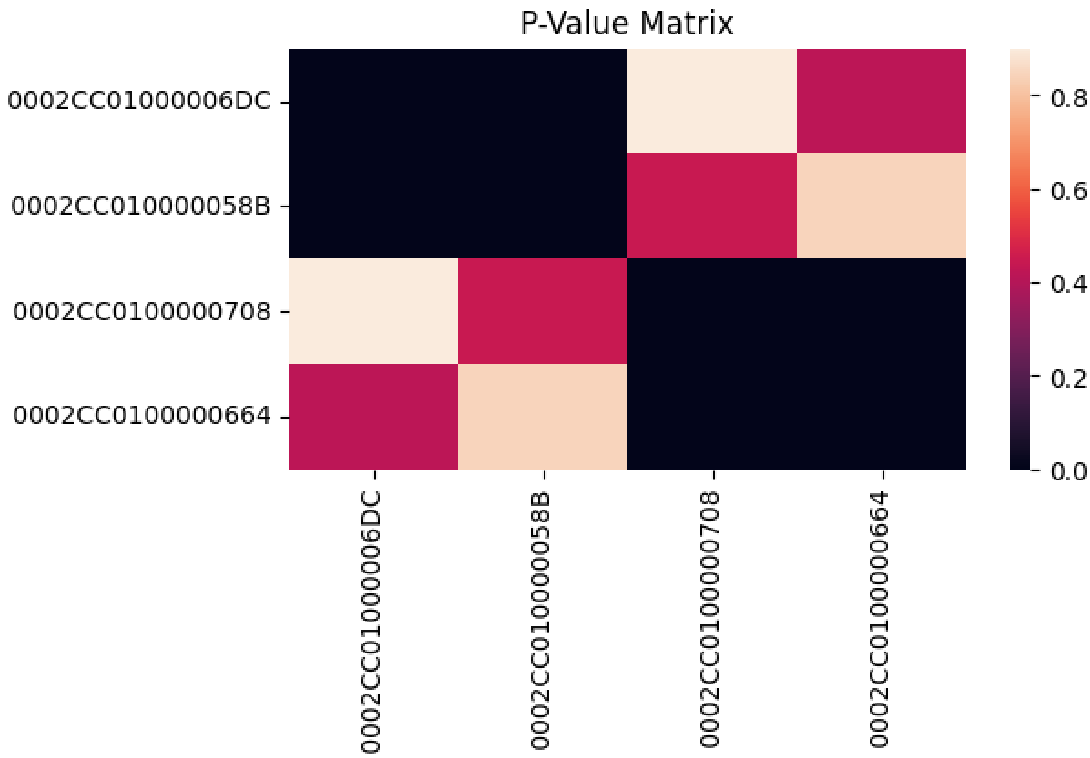
[14]:

	0002CC01000006DC	0002CC010000058B	0002CC0100000708
0002CC01000006DC	NaN	0.000494	0.900473
0002CC010000058B	0.000494	NaN	0.444846
0002CC0100000708	0.900473	0.444846	NaN
0002CC0100000664	0.417092	0.844769	0.0

	0002CC0100000664
0002CC01000006DC	0.417092
0002CC010000058B	0.844769
0002CC0100000708	0.0
0002CC0100000664	NaN

[15]:
```
plt.figure(figsize=(6,3))
sns.heatmap(p_values.fillna(0))
plt.title("P-Value Matrix")
```

[15]: Text(0.5, 1.0, 'P-Value Matrix')



[16]:

```
alpha = 0.05
for i in detected_devices:
    for j in detected_devices:
        if i != j and p_values.loc[i, j] < alpha:
            print(f"Significant correlation between {i} and {j}: p-value =␣
↪{p_values.loc[i, j]}")
        else:
            print(f"No significant correlation between {i} and {j}: p-value =␣
↪{p_values.loc[i, j]}")
```

```
No significant correlation between  0002CC01000006DC and  0002CC01000006DC:
p-value = nan
Significant correlation between  0002CC01000006DC and  0002CC010000058B: p-value
= 0.00049444933460338 82
No significant correlation between  0002CC01000006DC and  0002CC0100000708:
p-value = 0.9004725513440803
No significant correlation between  0002CC01000006DC and  0002CC0100000664:
p-value = 0.4170919624103882
Significant correlation between  0002CC010000058B and  0002CC01000006DC: p-value
= 0.00049444933460338 82
No significant correlation between  0002CC010000058B and  0002CC010000058B:
p-value = nan
```

```
No significant correlation between  0002CC010000058B and  0002CC0100000708:
p-value = 0.44484577190174046
No significant correlation between  0002CC010000058B and  0002CC0100000664:
p-value = 0.8447688502485741
No significant correlation between  0002CC0100000708 and  0002CC01000006DC:
p-value = 0.9004725513440803
No significant correlation between  0002CC0100000708 and  0002CC010000058B:
p-value = 0.44484577190174046
No significant correlation between  0002CC0100000708 and  0002CC0100000708:
p-value = nan
Significant correlation between  0002CC0100000708 and  0002CC0100000664: p-value
= 5.185484086004395e-09
No significant correlation between  0002CC0100000664 and  0002CC01000006DC:
p-value = 0.4170919624103882
No significant correlation between  0002CC0100000664 and  0002CC010000058B:
p-value = 0.8447688502485741
Significant correlation between  0002CC0100000664 and  0002CC0100000708: p-value
= 5.185484086004395e-09
No significant correlation between  0002CC0100000664 and  0002CC0100000664:
p-value = nan
```

Significant correlations is found between the following pairs:

> 0002CC01000006DC and 0002CC010000058B: p-value = 0.0005 0002CC0100000708 and 0002CC0100000664: p-value = 5.19e-09 0002CC0100000664 and 0002CC0100000708: p-value = 5.19e-09

These results indicate strong statistical associations between the activity levels of these device pairs, suggesting that activity in one subject could influence or be associated with activity in another.

Overall, the significant correlations suggest that some subjects' activity levels are closely related, which might indicate a pattern or influence between these subjects. However, the absence of significant correlations for most pairs means that not all subjects' activities are related, highlighting a mixed pattern of influence across subjects.

Only a few subjects shown a significant correlations, there is no evidence suggesting that the activity levels of one subject can influence or be associated with those of another. Therefore, we fail to reject the null hypothesis, indicating that no correlation between the activity levels of one subject and the activity levels of other subjects.

[ ]:

```

```