Analysis of electricity demand of five states in Australia

Import packages

In [1]:
```
#import pac:
import pandas as pd
import plotly.graph_objects as go
import plotly.express as px
import numpy as np
import plotly.figure_factory as ff
import folium
from plotly.subplots import make_subplots
```

Data cleaning and Data Preparation

In [2]:
```
df_elec = pd.read_csv("C:\\Users\\acer\\OneDrive\\DataDisca\\16_10_2021 Task 01\\electricity1.txt", header = None)
```

```
df_elec
```

In [3]:
```
df_elec = df_elec.transpose()
df_elec
```

In [4]:
```
df_elec.to_csv('C:\\Users\\acer\\OneDrive\\DataDisca\\16_10_2021 Task 01\\electricity.csv')
```

In [5]:
```
df_elec = pd.read_csv("C:\\Users\\acer\\OneDrive\\DataDisca\\16_10_2021 Task 01\\df_elec.csv")
df_elec
```

In [6]:
```
df_elec.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 230735 entries, 0 to 230734
Data columns (total 5 columns):
 #   Column  Non-Null Count   Dtype  
---  ------  --------------   -----  
 0   NSW     230735 non-null  float64
 1   VIC     230735 non-null  float64
 2   QUN     230735 non-null  float64
 3   SA      230735 non-null  float64
 4   TAS     230735 non-null  float64
dtypes: float64(5)
memory usage: 8.8 MB
```

In [7]:
```
df_elec.describe()
```

Out[7]:

	NSW	VIC	QUN	SA	TAS
count	230735.000000	230735.000000	230735.000000	230735.000000	230735.000000
mean	6740.520038	4638.967741	4315.976740	1290.663312	507.201424
std	1361.919627	836.368163	875.444258	301.107272	147.323078
min	3498.385270	2688.516606	2008.623448	488.835380	-233.906816
25%	5751.777576	3975.544104	3601.204502	1063.574139	395.649342
50%	6783.570728	4573.126086	4368.535136	1271.670638	489.643232
75%	7658.946062	5233.423854	4947.085684	1459.678114	594.000053
max	12865.795820	9494.010992	7514.436522	3182.476646	1093.502130

.

Summary Table

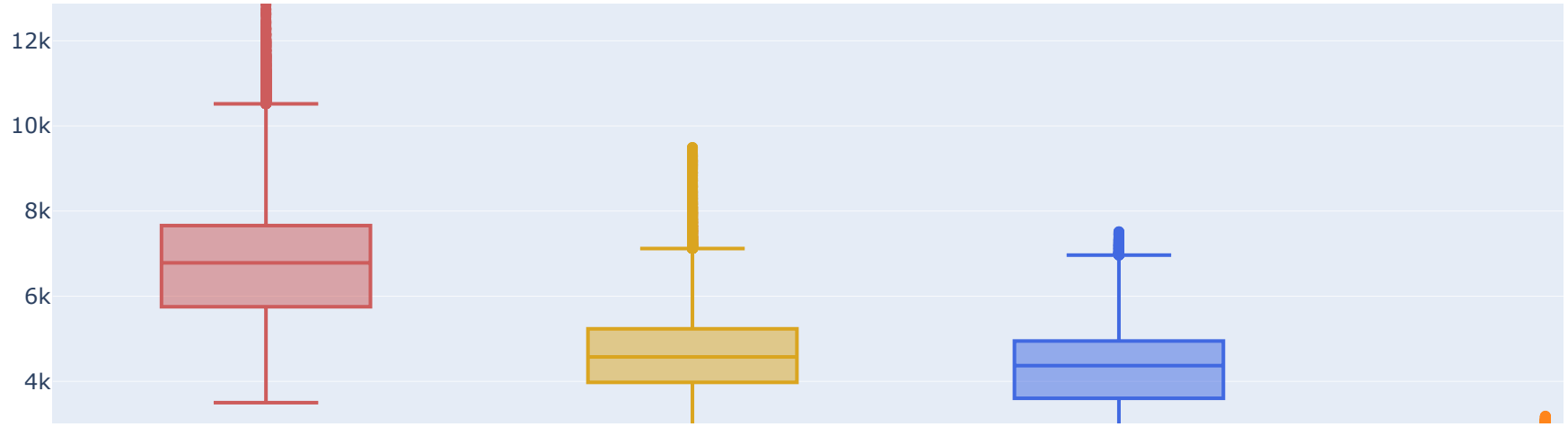
Field Name	Description	Pandas Data Type	Data Scale	Max Value	Min Value	Missing Value Count
NSW	New South Wales	float64	230735	3498.385270	3498.385270	non
VIC	Victoria	float64	230735	9494.010992	2688.516606	non
QUN	Queensland	float64	230735	7514.436522	2008.623448	non
SA	South Australia	float64	230735	3182.476646	488.835380	non
TAS	Tasmania	float64	230735	1093.502130	-233.906816	non

Data set has five (5) column and two hundred thirty thousand seven hundred and thirty-five (230,735) rows. Also data set has lot of unique values.

.

Visualization

In [8]:
```
fig = go.Figure()
fig.add_trace(go.Box(y=df_elec.NSW, name='NSW', marker_color = 'indianred'))
fig.add_trace(go.Box(y=df_elec.VIC, name = 'VIC', marker_color = 'goldenrod'))
fig.add_trace(go.Box(y=df_elec.QUN, name='QUN', marker_color = 'royalblue'))
fig.add_trace(go.Box(y=df_elec.SA, name = 'SA', marker_color = '#FF851B'))
fig.add_trace(go.Box(y=df_elec.TAS, name = 'TAS', marker_color = '#3D9970'))
fig.show()
```



The descriptive statistics are graphically visualized by Above boxplots.
Clearly show by box plots that NSW and TAS states have the highest and lowest average demand respectively and VIC and QUN states's average demanda are nearly close.

In [9]:
```
colors = ['indianred', 'goldenrod', 'royalblue', '#FF851B','#3D9970']
fig1 = px.line(df_elec, labels=dict(value = "Demand", index ="Time (half hour)"), color_discrete_sequence= colors)
fig1.show()
```