Living By The Sea. Should I?
- weather data analysis

Magdalene Cheong
Feb 2022

Introduction

I have been mulling over this for some time now: where shall I move to? Perhaps it is not such a big deal since Singapore is such a small country; there is not a huge difference from town to town. Proximity to the Central Business District at the south of the island is one of the main considerations for many but holds no importance to me. I dream of a place that is away from the hustling and bustling of city living.

Being a small city-state, there is not a lot of options for a quiet life. Living near the sea draws me. I imagine: facing the sea with my eyes closed after a long day's work, I listen to the murmurs of the sea and stand captivated by a conversation I do not understand; the gentle cool breeze brushes me by and carries with it all the cares of the day.

Oh but is it warm damp air blowing against my face? Will humidity be higher near the coast? It's supposed to be cooler near the sea, right? Alright, hold on to the cheque first.

One might argue that sitting on the equator, Singapore will hardly notice any temperature difference across the island, and that even if there is, difference of a few degrees will not matter. Perhaps? In a tropical country as Singapore, I would surely go for the coolest option possible.

My curiosity is piqued. I need to satisfy this curiosity: is it indeed cooler near the coast?

Should I move near the coast?

Contents

Objectives

To compare weather conditions near the coast against inland. 3 questions:

- How hot can it get during the hottest time of the day?
- How does the temperature near the coast differ from inland during evening when I would often be outdoors?
- Will it be colder late at night?

Time period of concern:

- Day time period
  - Defined by the hottest period over the day. This period coincides with high UV Index. Use UV Index to find the cut off time.
- Evening time period
  - Defined by the typical time most people would spend on recreation/outdoor after work and before bedtime.
- Night time period
  - Defined by typical bedtime

Objectives

Parameters:

- Main parameter for comparison: air temperature.
- Other parameters to help with understanding: humidity, rainfall, wind speed and wind direction.

Side note:

The following data extraction, preparation and data cleaning were not limited to only the locations of interest but covers all other stations. This is to have a complete dataset that can also be used for other analysis.

Limitations

- Housing in Singapore is mostly in high rise buildings. The measurement equipment by the Meteorological Service Singapore (MSS), a division under National Environment Agency (NEA) are placed nearer ground level in an open area. While precautions are taken to ensure accurate measurements such that air temperature readings are not affected by ground heat etc (Info on weather stations), the exact weather conditions experienced at high-rise buildings may not be the same depending on the facing and the architectural design.
- Parameters like wind speed and wind directions data are not strong indicators as the experience is dependent on the unit facing and if it is blocked by other buildings.

Brief Description of Dataset

- Source: https://data.gov.sg/developer
  - https://api.data.gov.sg/v1/environment/air-temperature
  - https://api.data.gov.sg/v1/environment/relative-humidity
  - https://api.data.gov.sg/v1/environment/rainfall
  - https://api.data.gov.sg/v1/environment/wind-speed
  - https://api.data.gov.sg/v1/environment/wind-direction
  - https://api.data.gov.sg/v1/environment/uv-index
- The dataset is related to weather measurements collected by metrological stations over Singapore.
- Parameters extracted: air temperature, relative humidity, rainfall, wind speed, wind direction and UV Index readings.
- The dataset consists of data from 20 stations. Geographical coordinates of the measurement stations are also extracted.
- Data is extracted for the time period of 5 years from 1 Jan 2017 to 31 Dec 2021 daily.
- Time interval between data points differs between parameters and hourly data is extracted for all.
- UV Index readings are available between 7am to 7pm.

Data Extraction, Preparation & Cleaning

- Prepare date list and time list.
  - Generate a date list to cover dates over intended period.
  - Generate a hourly time list in a format that facilitates extracting data from the APIs:

Hourly time list generation

```
1 time_lst = ["T0"+str(x)+":01" for x in range(10)] + ["T"+str(x)+":01" for x in range(10,24)]
2
3 print(time_lst)
```

```
['T00:01', 'T01:01', 'T02:01', 'T03:01', 'T04:01', 'T05:01', 'T06:01', 'T07:01', 'T08:01', 'T09:01', 'T10:01', 'T11:01', 'T12:0
1', 'T13:01', 'T14:01', 'T15:01', 'T16:01', 'T17:01', 'T18:01', 'T19:01', 'T20:01', 'T21:01', 'T22:01', 'T23:01']
```

  - Sample of timestamp on raw data: 2020-05-05T23:01:00+08:00
  - To ensure that the data is correctly extracted, the letter ‘T’ is included in generating the time list. This prevents confusion between hour, minute and second and thus also prevents dirty data.

Data Extraction, Preparation & Cleaning

- Write Function to extract data from URL.
  - In order capture missing data due to glitches or other reasons, dates of which data failed to get extracted are immediately returned and fed back into code to re-extract. This prevents missing data.

Get data from URL

```
def get_data(url, date_lst):

    hr_data_lst = []
    no_data = []

    for i in date_lst:
        params['date'] = i
        try:
            resp = requests.get(url,params)
            resp_py = resp.json()
            data = resp_py['items']
            for item in data:
                for t in time_lst:
                    if t == item['timestamp'][10:16]:
                        hr_data_lst.append( {'timestamp':item['timestamp'], 'readings':item['readings']} )
        except:
            no_data.append(i)

    print("Unsuccessful dates: ", no_data)
    print(len(hr_data_lst))
    return hr_data_lst
```

Data Extraction, Preparation & Cleaning

- Write Function to format data into usable format.
  - The data extracted from the URL consists of list of dictionaries. This function flattens the data so that it can be loaded into pandas dataframe for analysis. This ensures that all data extracted across all features are in the same format.
  - Similarly, data that are not successfully extracted is returned.

Clean data into usable format

```
def get_clean_data(hr_data_lst):
    data_lst = []
    data_reading = []
    rec = []

    for i in hr_data_lst:
        try:
            for j in i['readings']:
                data_reading.append( { 'timestamp':i['timestamp'], 'station_id':j['station_id'], label : j['value']} )
            data_lst = data_lst + data_reading
            data_reading = []
        except:
            rec.append(i)

    print("Unsuccessful clean data: ", rec)
    return data_lst

```

Data Extraction, Preparation & Cleaning

- Data Extraction:
  - ‘label’ is extracted from the URL and identifies the parameter in that URL (e.g.wind-direction)
  - ‘label’ is then used in the function get_clean_data() to rename the column for readings.
  - ‘label’ is also used to concatenate into the filename before saving as csv file.
  - First and last date of a particular are recorded as part of the file name. These files are then combined in python into 1 csv file per parameter.

Enter URL

```
url = input("Enter url: ")

params = {}
label = url.split('/')[-1]
print(label)
```

```
Enter url: https://api.data.gov.sg/v1/environment/wind-direction
wind-direction
```

```
results = get_clean_data(get_data(url, date_lst_gen(first_date, last_date)))
```

```
df = pd.DataFrame(results)
```

```
filename = label + "_" + str(first_date) + "_" + str(last_date) + ".csv"
df.to_csv(filename)
print("Saved!")
```

Data Extraction, Preparation & Cleaning

- Unnecessary columns are dropped.
  - Air-temperature data as sample:

```
1 df_temp=pd.read_csv("air-temp.csv")
```

```
1 df_temp.drop(columns="Unnamed: 0", inplace=True)
```

```
1 df_temp
```

	timestamp	station_id	air-temperature
0	2017-01-01T00:01:59+08:00	S109	26.0
1	2017-01-01T00:01:59+08:00	S117	26.9
2	2017-01-01T00:01:59+08:00	S107	27.0
3	2017-01-01T00:01:59+08:00	S43	26.9
4	2017-01-01T00:01:59+08:00	S108	27.0
...	...	...	...
633780	2021-07-21T23:01:00+08:00	S122	29.2
633781	2021-07-21T23:01:00+08:00	S115	29.5
633782	2021-07-21T23:01:00+08:00	S24	29.1
633783	2021-07-21T23:01:00+08:00	S116	29.3
633784	2021-07-21T23:01:00+08:00	S100	28.9

633785 rows × 3 columns

Data Extraction, Preparation & Cleaning

- Check for null and data type.
  - As shown, some stations returns a huge number of nulls for temperature data.

```
1 temp = pd.pivot(df_temp, values="air-temperature", index=["timestamp"], columns=["station_id"])
```

```
1 temp
```

station_id	S06	S100	S102	S104	S106	S107	S108	S109	S111	S115	...	S117	S121	S122	S24	S24B	S43	S44	S50	S60	S96
timestamp																					
2017-01-01T00:01:59+08:00	26.2	26.3	27.3	25.9	25.1	27.0	27.0	26.0	NaN	26.7	...	26.9	26.0	25.5	25.9	26.4	26.9	25.5	NaN	NaN	26.6
2017-01-01T01:01:59+08:00	26.3	26.4	27.1	25.9	25.0	26.9	27.0	25.9	NaN	26.5	...	26.5	25.8	26.1	25.9	26.3	27.0	25.3	NaN	NaN	26.7
2017-01-01T02:01:59+08:00	26.2	26.3	27.1	25.7	25.2	26.8	26.9	26.1	NaN	26.2	...	26.5	25.7	26.0	25.8	26.1	26.7	25.1	NaN	NaN	26.5
2017-01-01T03:01:59+08:00	26.2	26.0	26.9	25.6	25.5	26.8	26.5	25.9	NaN	25.9	...	26.3	25.5	26.0	25.7	26.1	26.5	24.8	NaN	NaN	26.2
2017-01-01T04:01:59+08:00	26.1	25.9	26.7	25.5	25.4	26.6	26.3	25.9	NaN	25.6	...	26.0	25.3	25.6	25.7	26.1	26.5	24.6	NaN	NaN	26.1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2021-12-31T19:01:00+08:00	NaN	25.0	NaN	24.8	25.6	26.2	25.4	25.2	NaN	24.8	...	NaN	24.8	NaN	25.9	NaN	25.6	24.3	24.8	NaN	NaN
2021-12-31T20:01:00+08:00	NaN	24.5	NaN	23.8	24.4	26.4	25.8	24.2	NaN	25.1	...	NaN	24.8	NaN	25.4	NaN	25.6	24.2	24.7	NaN	NaN
2021-12-31T21:01:00+08:00	NaN	24.3	NaN	24.3	NaN	25.5	24.9	23.4	NaN	25.1	...	NaN	24.7	NaN	24.7	NaN	24.6	24.1	23.9	NaN	NaN
2021-12-31T22:01:00+08:00	NaN	24.5	NaN	24.4	NaN	25.7	24.5	24.1	NaN	25.2	...	NaN	24.3	NaN	24.8	NaN	24.5	24.3	23.1	NaN	NaN
2021-12-31T23:01:00+08:00	NaN	24.5	NaN	25.3	NaN	25.5	24.6	24.4	NaN	25.3	...	NaN	24.1	NaN	24.6	NaN	24.9	24.1	23.6	NaN	NaN

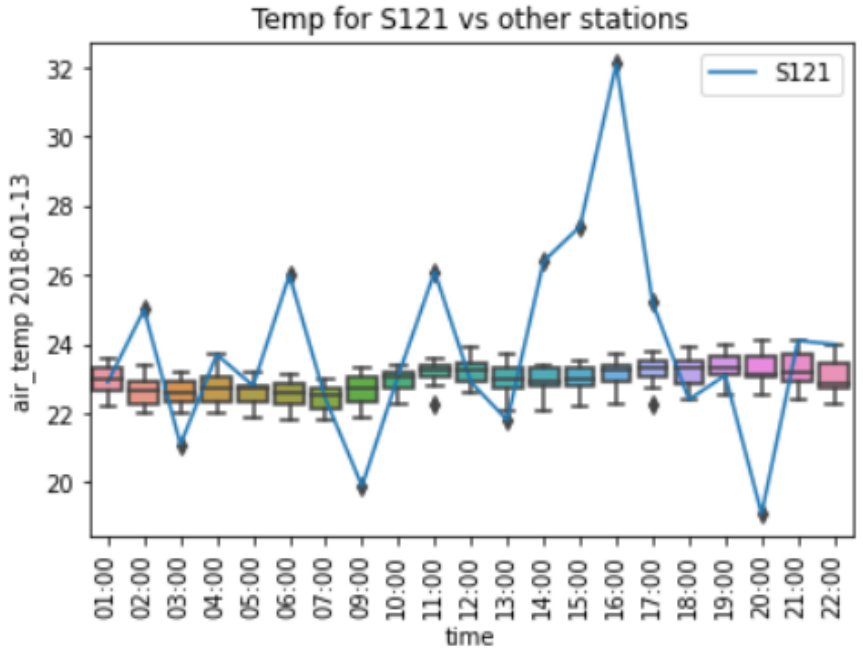
41588 rows × 21 columns

```
1 temp.isna().sum()
```

```
station_id
S06     39537
S100     1896
S102    34666
S104     4706
S106     6021
S107     1160
S108     7049
S109     3960
S111     7640
S115     3178
S116     1832
S117    10727
S121     2500
S122    10885
S24       547
S24B    38867
S43       997
S44      2513
S50      4969
S60     16882
S96     39031
dtype: int64
```

Data Extraction, Preparation & Cleaning

- In order to have a fair dataset, data extracted for all parameters are combined by SQL inner join to ensure there is data for every parameter at every timestamp across the stations.
- Incomplete data from stations at specific times may also indicate instrument issues and is best to be removed from the dataset. One such example is encountered for station_id S121 on 2018-01-13. The readings from this instrument on that day are erratic and matches completely with the outliers from the boxplot (other stations). This set of data is eliminated after SQL join.



Data Extraction, Preparation & Cleaning

- Do a SQL inner join.
- Then remove duplicates with *drop_duplicates()*

```
query = '''
SELECT T.*, H.rel_humidity, R.rainfall, W.windspeed, WD.winddir
FROM df_temp as T
INNER JOIN df_hum as H
ON T.day = H.day and T.time = H.time and T.station_id = H.station_id
INNER JOIN df_rain as R
ON T.day = R.day and T.time = R.time and T.station_id = R.station_id
INNER JOIN df_wind as W
ON T.day = W.day and T.time = W.time and T.station_id = W.station_id
INNER JOIN df_wind_dir as WD
ON T.day = WD.day and T.time = WD.time and T.station_id = WD.station_id;
'''
```

```
col_temp = df_all.pop("air_temp")
df_all.insert(3, "air_temp", col_temp)
df_all
```

	station_id	day	time	air_temp	rel_humidity	rainfall	windspeed	winddir
0	S109	2017-01-01	00:00	26.0	90.6	0.0	2.1	24
1	S117	2017-01-01	00:00	26.9	82.1	0.0	2.2	347
2	S107	2017-01-01	00:00	27.0	89.8	0.0	1.3	2
3	S43	2017-01-01	00:00	26.9	88.1	0.0	2.2	36
4	S108	2017-01-01	00:00	27.0	87.3	0.0	11.5	356
...	...	...	...	...	...	...	...	...
496429	S122	2021-07-21	23:00	29.2	78.6	0.0	2.6	172
496430	S115	2021-07-21	23:00	29.5	75.3	0.0	4.3	61
496431	S24	2021-07-21	23:00	29.1	73.8	0.0	8.3	176
496432	S116	2021-07-21	23:00	29.3	86.7	0.0	6.3	153
496433	S100	2021-07-21	23:00	28.9	74.0	0.0	1.2	121

486731 rows × 8 columns

Data Extraction, Preparation & Cleaning

- Inspect data for outliers on the combined dataset.
- Min value (19.1˚C) for air_temp (temperature) is atypical of Singapore weather. Historically, temperature extremes like 19˚C did occur (19.4˚C on 30 Jan 1934 and 31 Jan 1934, and 19˚C on 14 Feb 1989).
- Min value (29.5%) for relative humidity is much smaller than value at 25th-percentile.
- Temperature and relative humidity has a high correlation.
- Rain is infrequent.

```
df_para = df[["air_temp", "rel_humidity", "rainfall", "windspeed", "winddir"]]
#df_para.head()
```

```
df_para.describe()
```

	air_temp	rel_humidity	rainfall	windspeed	winddir
count	486731.000000	486731.000000	486731.000000	486731.000000	486731.000000
mean	28.190119	79.493747	0.017505	4.945247	152.061443
std	2.062776	11.419369	0.221900	3.491800	111.570289
min	19.100000	29.500000	0.000000	0.000000	1.000000
25%	26.700000	72.200000	0.000000	2.500000	47.000000
50%	28.100000	81.200000	0.000000	4.300000	137.000000
75%	29.600000	87.900000	0.000000	6.500000	244.000000
max	35.900000	101.000000	12.600000	50.000000	359.000000

```
df_para.corr()
```

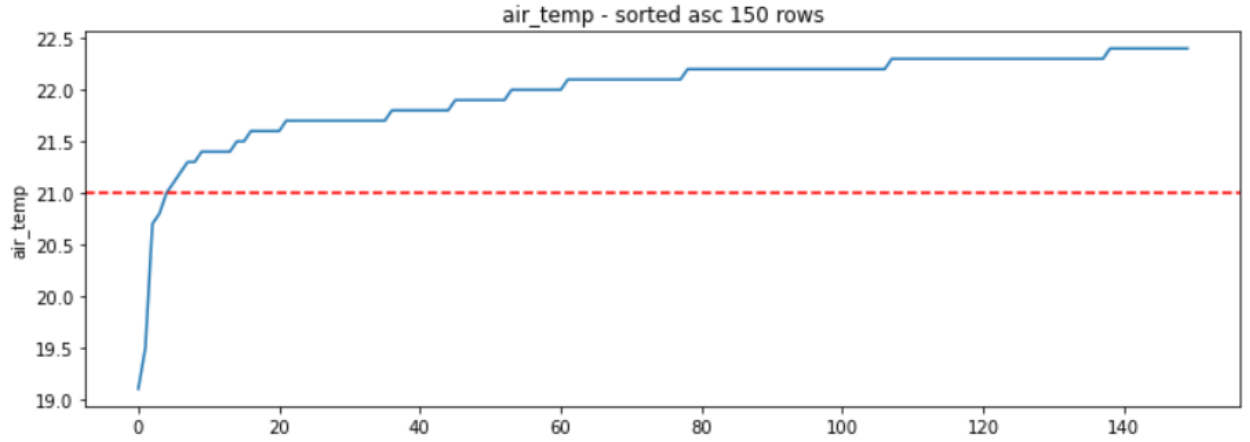
	air_temp	rel_humidity	rainfall	windspeed	winddir
air_temp	1.000000	-0.831098	-0.108698	0.161878	0.001320
rel_humidity	-0.831098	1.000000	0.089399	-0.162558	0.037593
rainfall	-0.108698	0.089399	1.000000	0.033695	0.028093
windspeed	0.161878	-0.162558	0.033695	1.000000	-0.107108
winddir	0.001320	0.037593	0.028093	-0.107108	1.000000

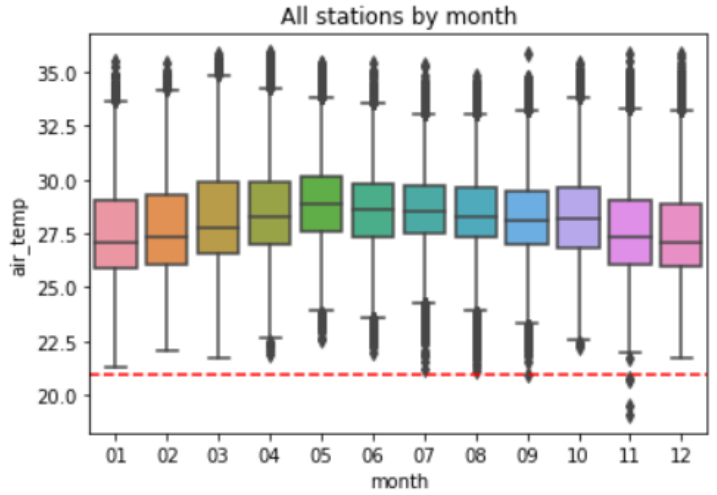
Data Extraction, Preparation & Cleaning

- Inspect data for outliers. Sort air_temp and plot first 150.
- A deep drop at the start of the plot. Temperature above 21˚C seems more representative, below which reading is atypical or measurement might be wrong.
- From boxplot (by month), 21˚C seems a reasonable cutoff for minimum temperature.

```
1 plt.figure(figsize=(12,4))
2 sns.lineplot(data=df_sorttemp["air_temp"][:150])
3 plt.axhline(21, c="red", linestyle= "dashed")
4 plt.title("air_temp - sorted asc 150 rows")
```

```
Text(0.5, 1.0, 'air_temp - sorted asc 150 rows')
```





Data Extraction, Preparation & Cleaning

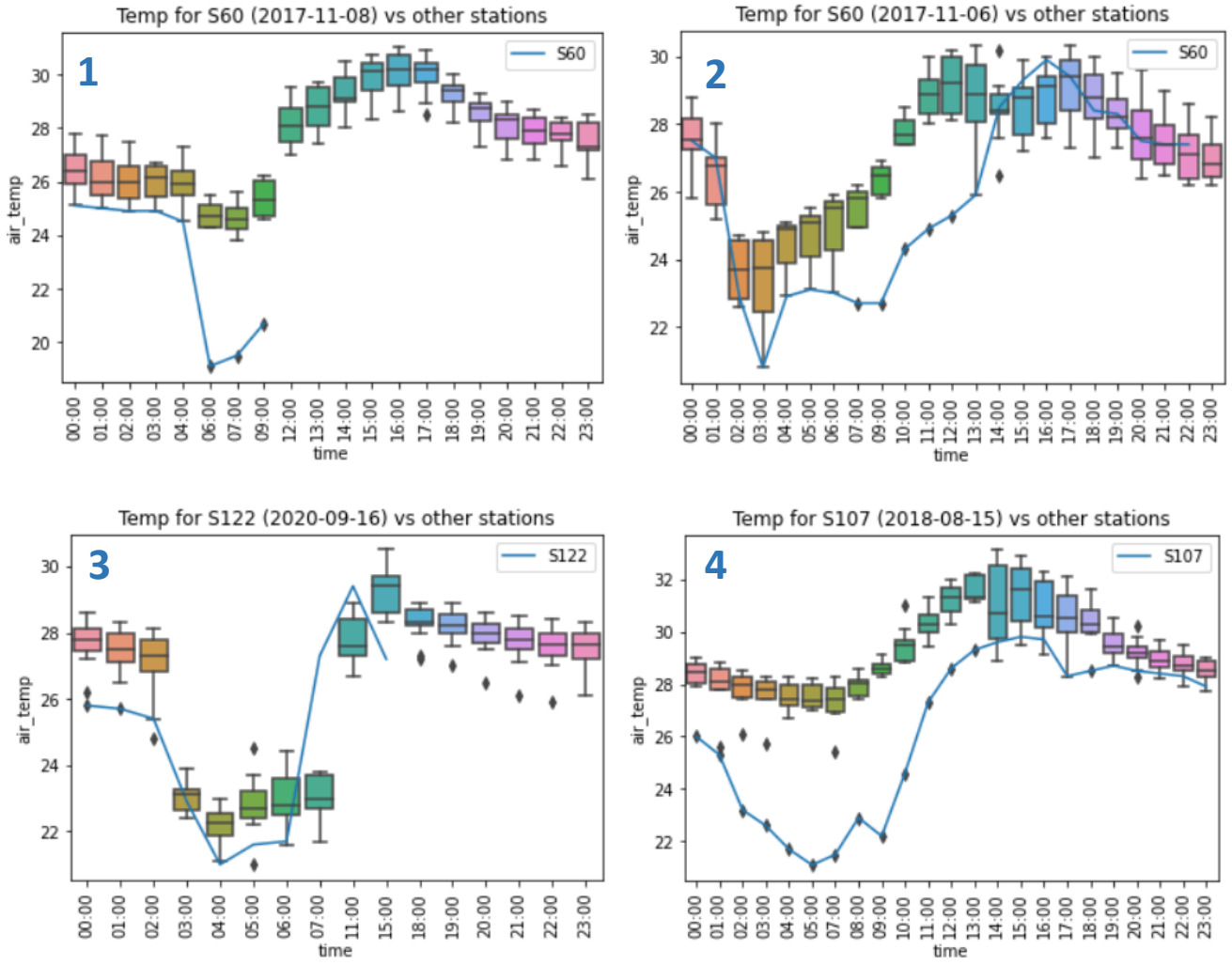
- From the table, all the readings less than 21˚C comes from station_id S60. Data for S60 (2017-11-08), S60 (2017-11-06), S122 (2020-09-16) and S107 (2019-08-15) should be investigated.
- S60 (Sentosa) and S107 (East Coast Parkway) are both near the sea.
- S122 is at Sembawang – inland.

```
1 df_sorttemp = df.sort_values(by=['air_temp'], ignore_index=True)
2 df_sorttemp.head(10)
```

	station_id	day	time	identifier	air_temp	humidity	rainfall	windspeed	winddir	name
0	S60	2017-11-08	06:00	2017-11-08_06:00_S60	19.1	94.8	0.0	7.00000	277	Sentosa
1	S60	2017-11-08	07:00	2017-11-08_07:00_S60	19.5	98.0	1.4	11.30000	199	Sentosa
2	S60	2017-11-08	09:00	2017-11-08_09:00_S60	20.7	85.6	0.0	9.50000	243	Sentosa
3	S60	2017-11-06	03:00	2017-11-06_03:00_S60	20.8	99.0	0.4	3.50000	233	Sentosa
4	S122	2020-09-16	05:00	2020-09-16_05:00_S122	21.0	59.8	0.2	1.60000	306	Sembawang Road
5	S107	2019-08-15	05:00	2019-08-15_05:00_S107	21.1	84.8	0.0	7.50000	128	East Coast Parkway
6	S50	2018-07-15	05:00	2018-07-15_05:00_S50	21.2	98.8	0.6	5.80000	322	Clementi Road
7	S104	2018-01-14	05:00	2018-01-14_05:00_S104	21.3	96.9	0.0	9.10001	6	Woodlands Avenue 9
8	S44	2018-08-08	01:00	2018-08-08_01:00_S44	21.3	97.1	0.0	11.60000	302	Nanyang Avenue
9	S24	2018-01-14	05:00	2018-01-14_05:00_S24	21.4	96.4	0.2	5.60000	315	Upper Changi Road North

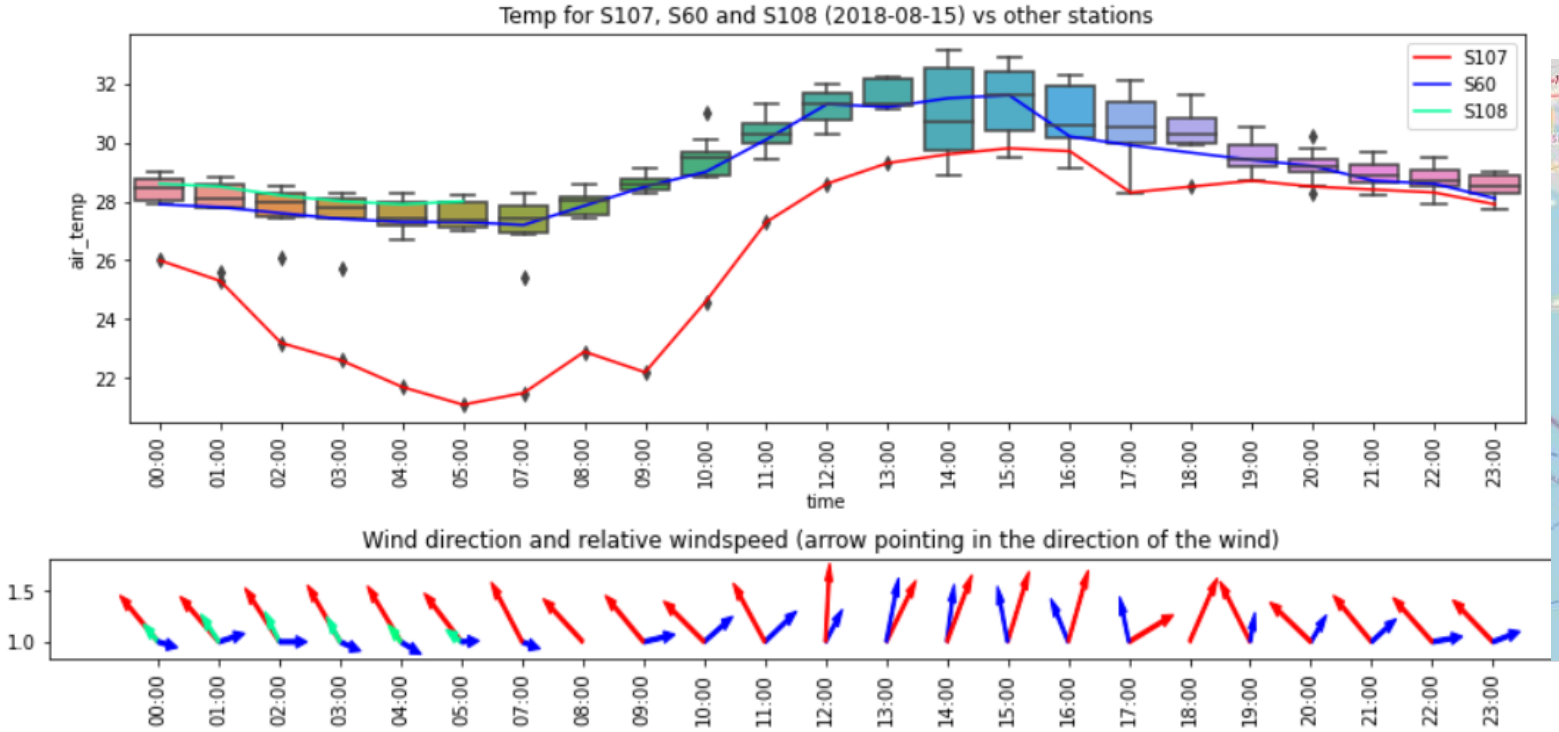
Data Extraction, Preparation & Cleaning

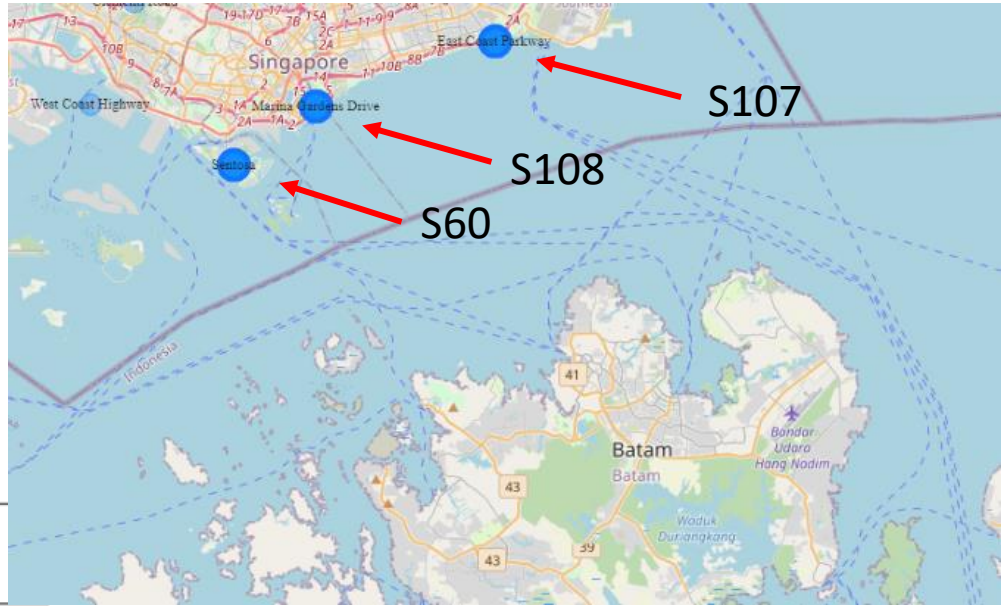
- Investigate air_temp outliers and data near 21˚C border.
- Graph 3 (S122 inland): There is an overall dip in temperature nation-wide. There is therefore unlikely measurement error.
- This gets really interesting at S60 & S107, both near the sea: can it really get that cold at the coastal areas or is it data error?
- Temperature at S60 and S107 converge with the other stations in the afternoon. That suggests the instruments were likely functional at that point of time. A faulty instrument is likely to behave like S121 as highlighted in Data Extraction section.
- The next page explores how temperature might interact with wind for stations near the coast (S107, S60 and S108).



Data Extraction, Preparation & Cleaning

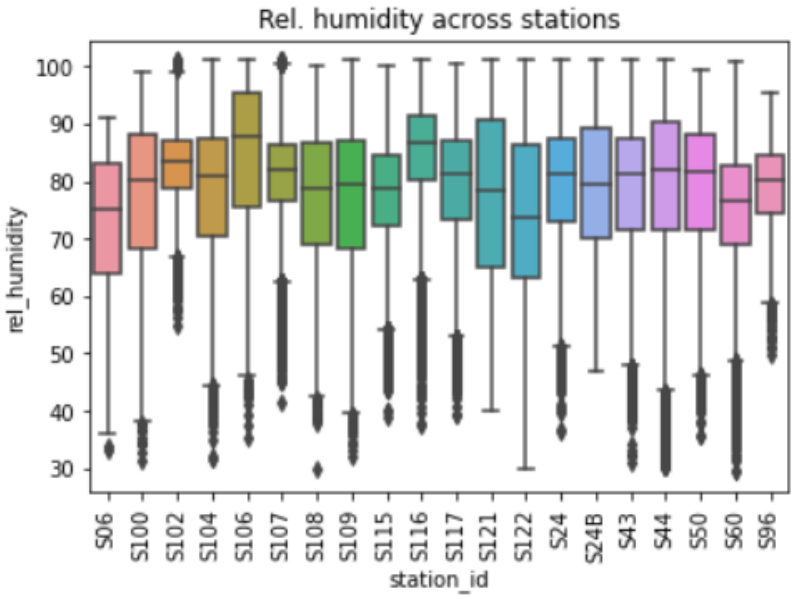
- Low temperature at S107 coincides with strong wind coming from the sea. Temperature difference between S107 and other stations narrowed from around noon (hottest time) but remained on the low side.
- Wind direction at S108 is similar to S107 but at much lower speed. S60 had wind at much lower speed and in a different direction in the time period S107 had low temperature. These could account for their higher temperature as compared to S107.
- Correlation coefficient between wind and temperature is very low (0.16). However, there is some indication of its influence here. Historically, 19˚C and 19.4 ˚C were experienced here in 1989 & 1934. Therefore, it is highly probable the low temperature data here are correct. Therefore, these data will be retained in this dataset.

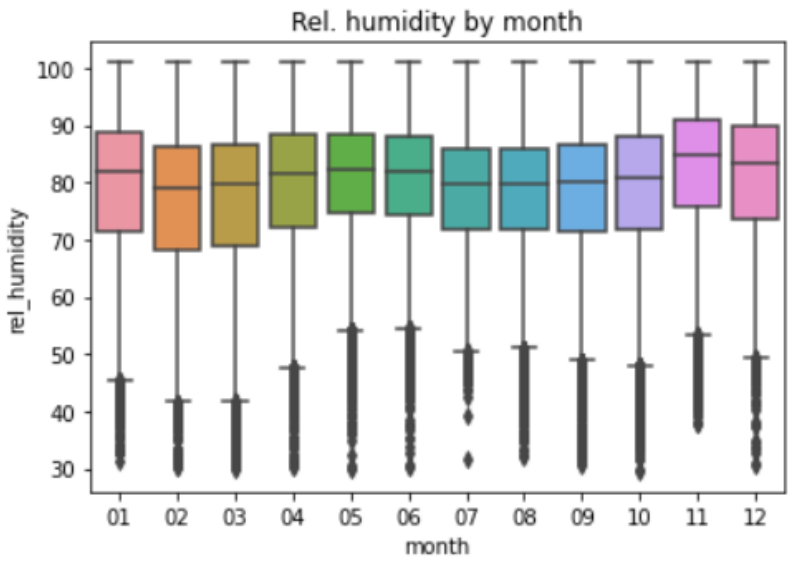


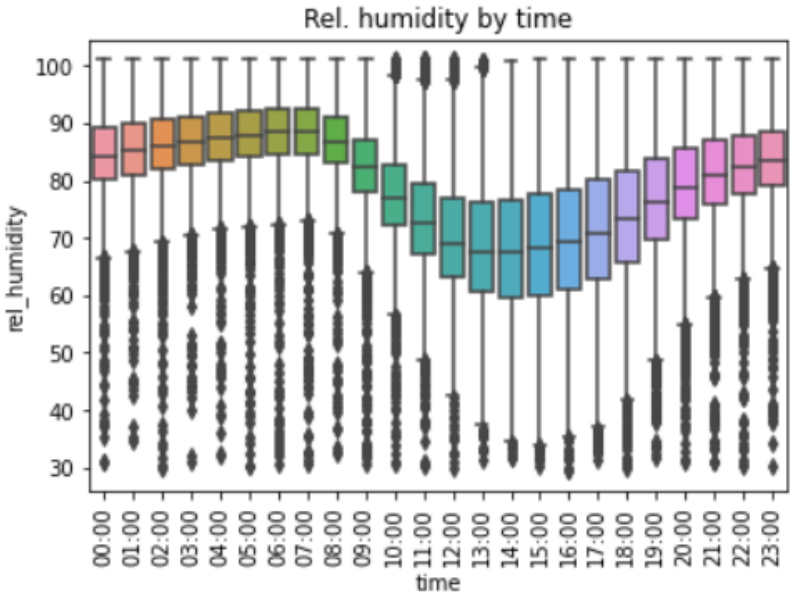


Data Extraction, Preparation & Cleaning

- Investigate if humidity below 72% is outlier.
- Relative humidity values below 25 percentile is not unique to any station or time period hence the probability of wrong measurement is low. No further investigation on low humidity will be done at this point.







Data Extraction, Preparation & Cleaning

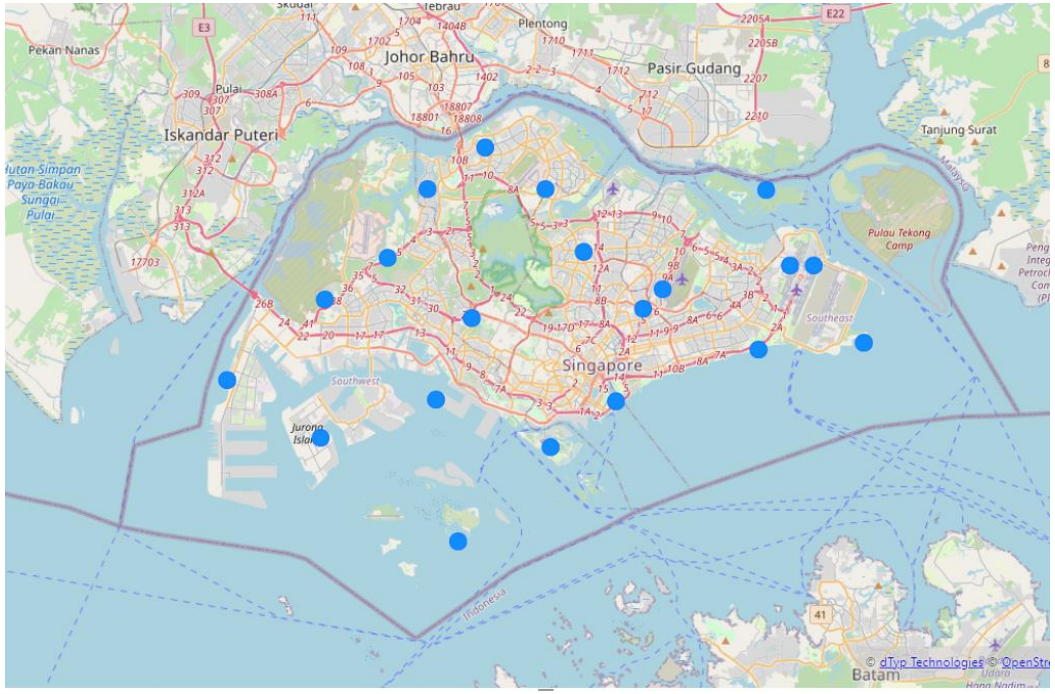
- Check sample size of each station_id. Some stations have much smaller sample size. Data for these stations should be dropped if analysis is at station level. However, if they are studied as a cluster (e.g. zonal), they may or may not be dropped.

```
1 df_all.groupby(['station_id']).size()
```

```
station_id
S06      2035
S100    38995
S102     7184
S104    35036
S106    28106
S107    38780
S108    22320
S109    33099
S115    37953
S116    38250
S117    30083
S121     1155
S122    14582
S24     40128
S24B      921
S43     39105
S44     30264
S50     32682
S60     23250
S96      2506
dtype: int64
```

Data Extraction, Preparation & Cleaning

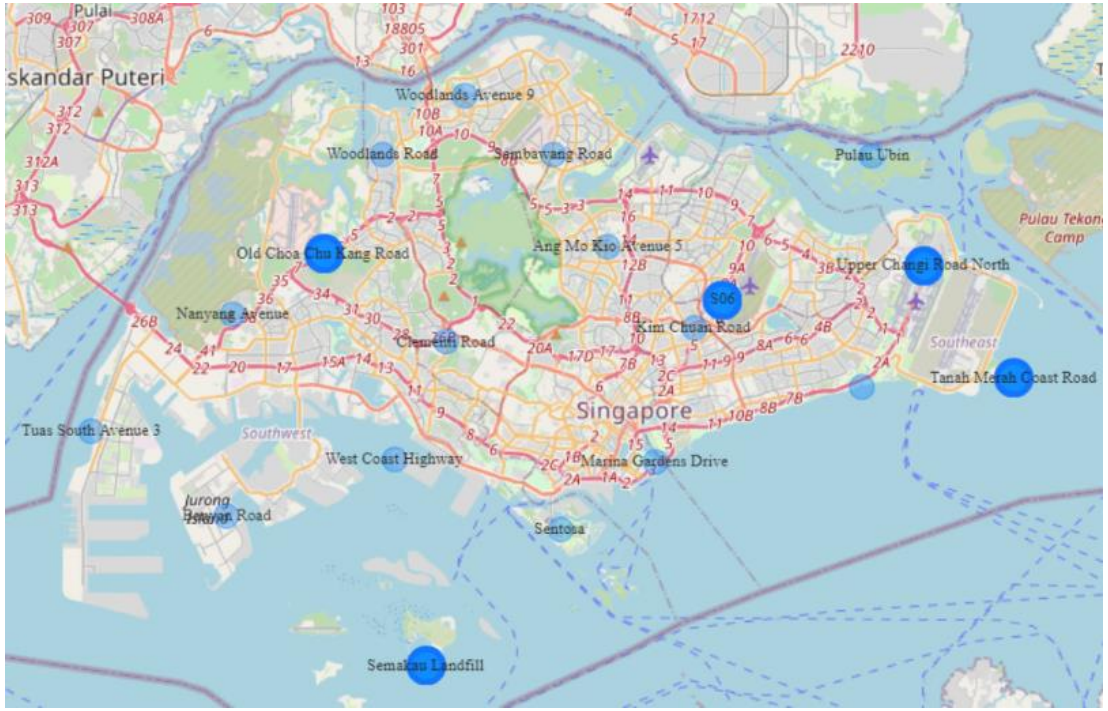
- Station_id and the geographical locations are extracted in the same way as the weather parameters.
- Plot done on MS PowerBI.



	station_id	name	long	lat
0	S109	Ang Mo Kio Avenue 5	103.84920	1.37640
1	S117	Banyan Road	103.67900	1.25600
2	S107	East Coast Parkway	103.96250	1.31350
3	S43	Kim Chuan Road	103.88780	1.33990
4	S108	Marina Gardens Drive	103.87030	1.27990
5	S44	Nanyang Avenue	103.68166	1.34583
6	S121	Old Choa Chu Kang Road	103.72244	1.37288
7	S106	Pulau Ubin	103.96730	1.41680
8	S06	S06	103.90070	1.35240
9	S102	Semakau Landfill	103.76800	1.18900
10	S122	Sembawang Road	103.82490	1.41731
11	S96	Tanah Merah Coast Road	104.03070	1.31750
12	S115	Tuas South Avenue 3	103.61843	1.29377
13	S24B	Upper Changi Road North	103.99800	1.36780
14	S24	Upper Changi Road North	103.98260	1.36780
15	S116	West Coast Highway	103.75400	1.28100
16	S104	Woodlands Avenue 9	103.78538	1.44387
17	S100	Woodlands Road	103.74855	1.41720
18	S50	Clementi Road	103.77680	1.33370
19	S60	Sentosa	103.82790	1.25000

Data Extraction, Preparation & Cleaning

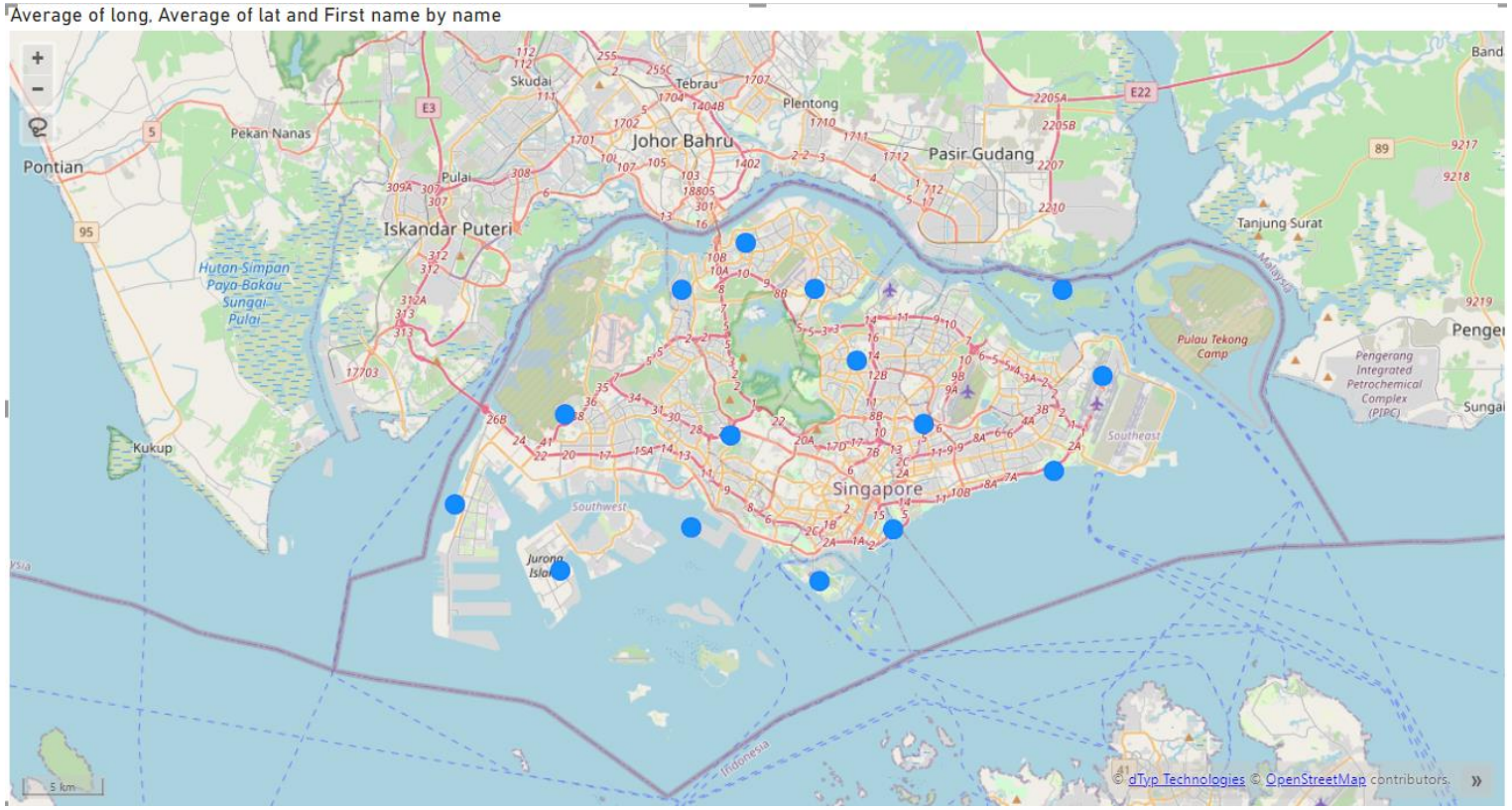
- Highlighted are stations with low sample size. These are non-residential areas with the exception of S24B.



	station_id	name	long	lat
0	S109	Ang Mo Kio Avenue 5	103.84920	1.37640
1	S117	Banyan Road	103.67900	1.25600
2	S107	East Coast Parkway	103.96250	1.31350
3	S43	Kim Chuan Road	103.88780	1.33990
4	S108	Marina Gardens Drive	103.87030	1.27990
5	S44	Nanyang Avenue	103.68166	1.34583
6	S121	Old Choa Chu Kang Road	103.72244	1.37288
7	S106	Pulau Ubin	103.96730	1.41680
8	S06	S06	103.90070	1.35240
9	S102	Semakau Landfill	103.76800	1.18900
10	S122	Sembawang Road	103.82490	1.41731
11	S96	Tanah Merah Coast Road	104.03070	1.31750
12	S115	Tuas South Avenue 3	103.61843	1.29377
13	S24B	Upper Changi Road North	103.99800	1.36780
14	S24	Upper Changi Road North	103.98260	1.36780
15	S116	West Coast Highway	103.75400	1.28100
16	S104	Woodlands Avenue 9	103.78538	1.44387
17	S100	Woodlands Road	103.74855	1.41720
18	S50	Clementi Road	103.77680	1.33370
19	S60	Sentosa	103.82790	1.25000

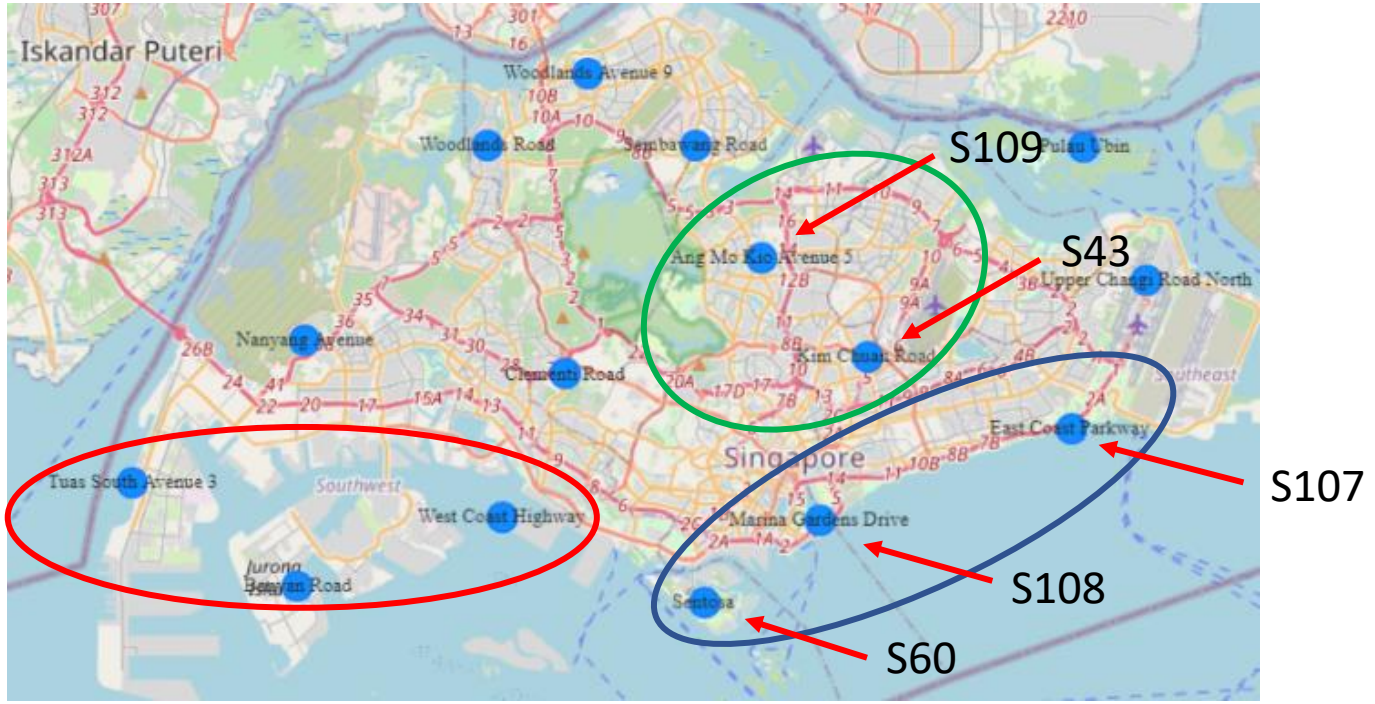
Data Extraction, Preparation & Cleaning

- Check distribution of stations after removing data from stations with low sample size. Distribution is even.



Define Scope of Data For Analysis

- Define location.
- Of the stations near the open sea, those in the red circle are near to heavy industries and do not appeal for residence personally.
- Those in the blue consists of residence ranging from public housing to private homes to premium homes in Sentosa. House pricing at Sentosa is exorbitant but this location will be compared for curiosity's sake.
- It is not possible to get the weather stations exactly where the current or preferred residence is, therefore the nearest possible are used (circled in green).



Define Scope of Data For Analysis

- Define time period for Test 1 (compare hottest time period).
- Use UV index to define the time period over the day where temperature data should be used for comparison.
- According to the chart below, UV index 6 and above is rated as ‘High’. On the UV chart on the right, UV index 6 corresponds to time period between 11am to 4pm.
- Cross referencing to temperature chart below it, 11am to 4pm is also the warmest time period over the day.
- Since most values at 11am and 4pm fall below UV index 6, the time period to use should thus be from 12pm to 3pm.

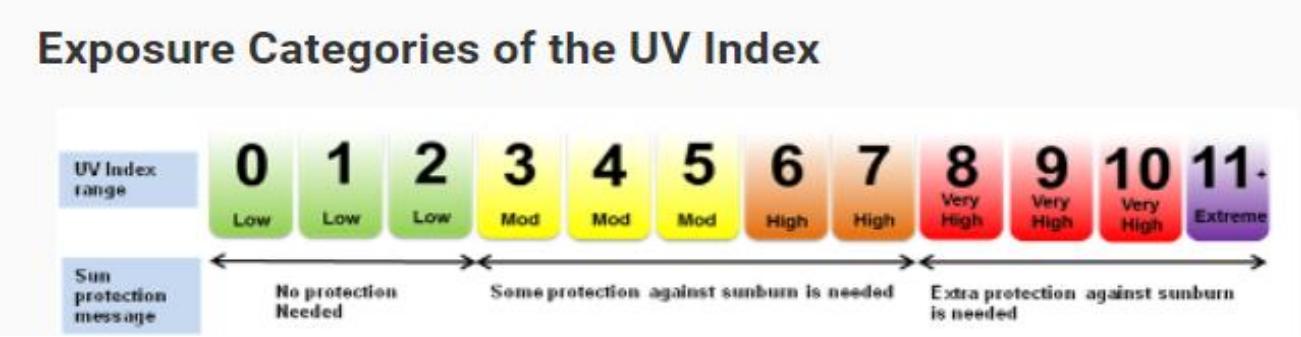
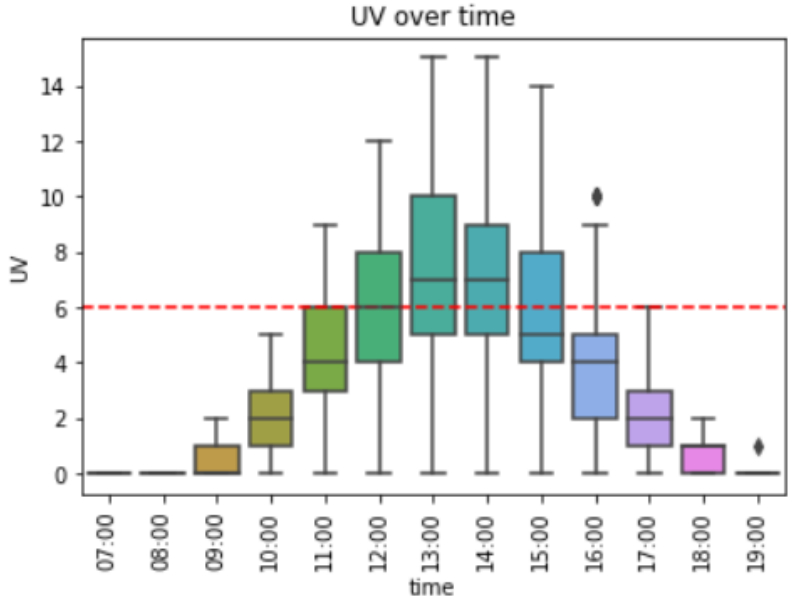
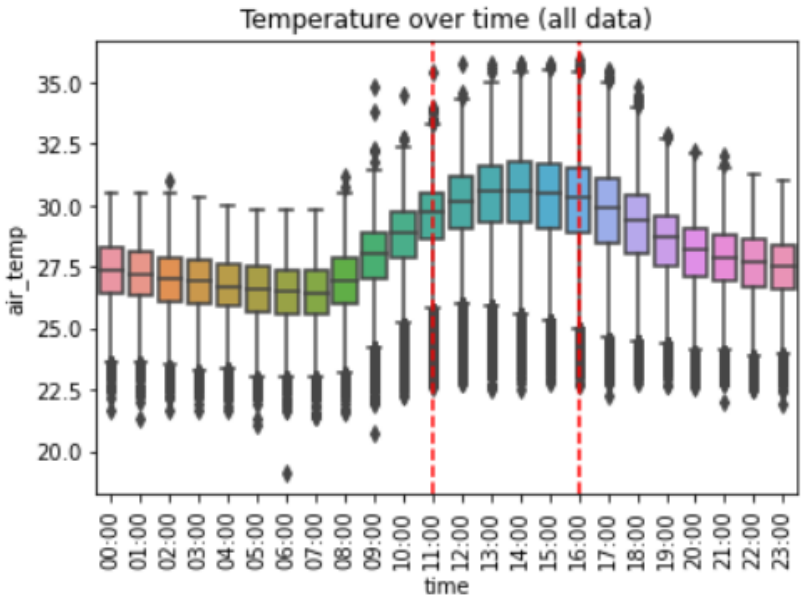


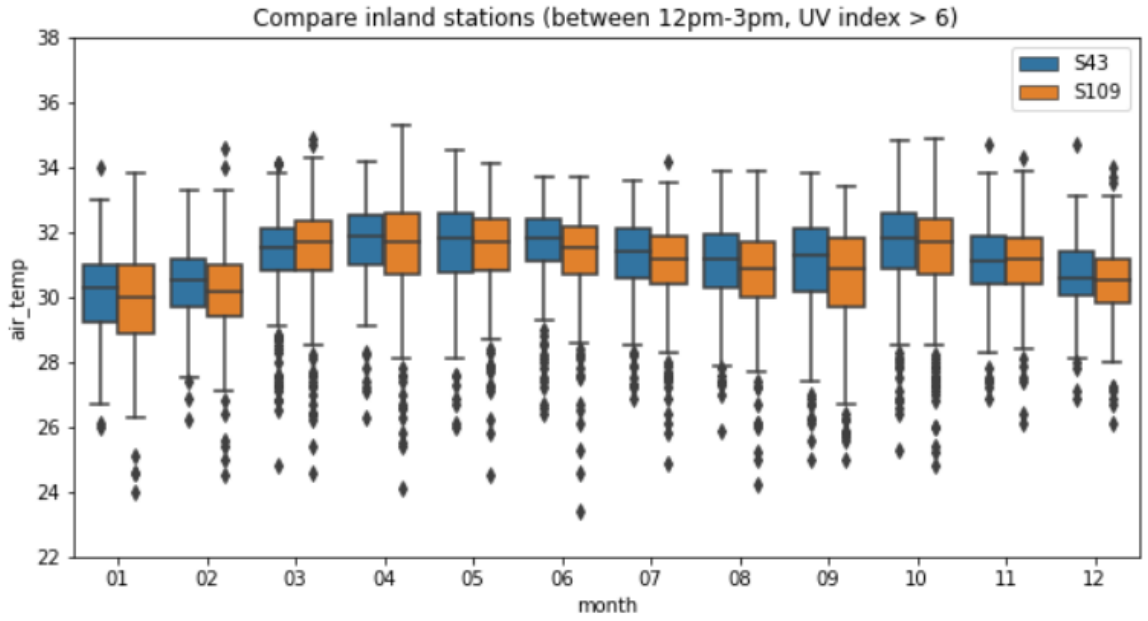
Chart from https://www.nea.gov.sg/weather/ultraviolet-index/uv-radiation-uv-index

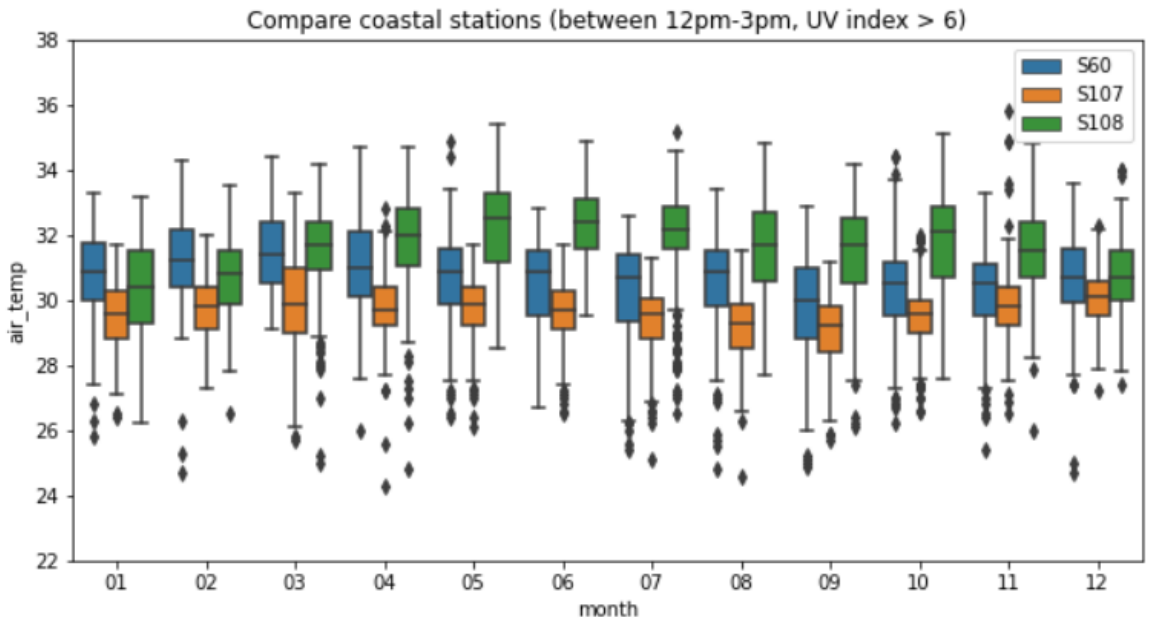




Define Scope of Data For Analysis

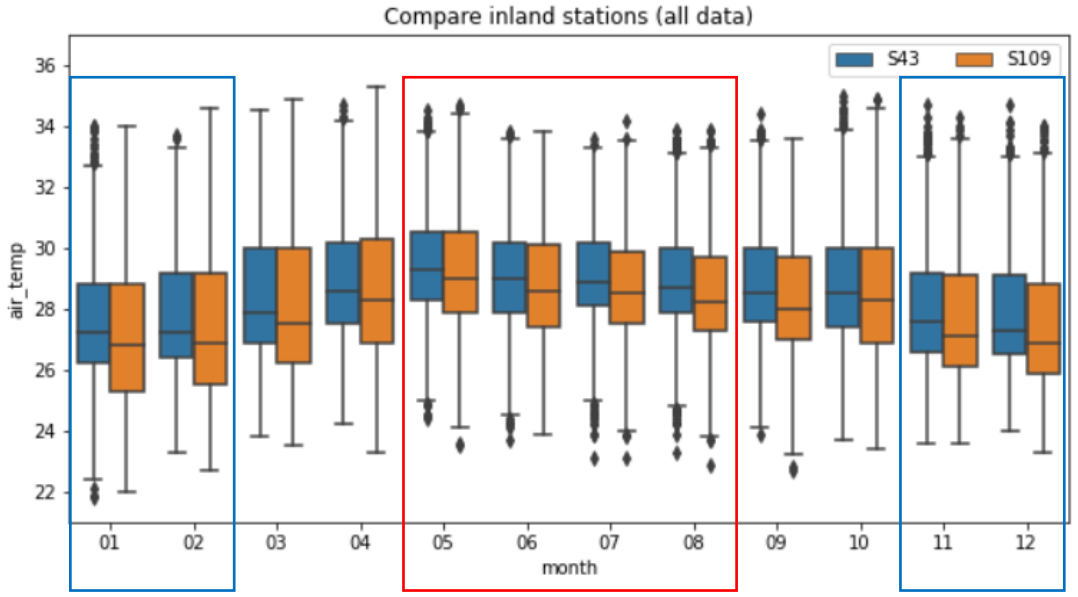
- Define grouping.
- Charts here are based on hottest time of the day - temperature between 12-3pm and where UV index > 6.
- Temperature is comparable between both stations for inland group but vastly differ for coastal group.
- Each of the weather station near the coast - S60, S107 and S108 - will be individually compared against Inland group (S43+S109)

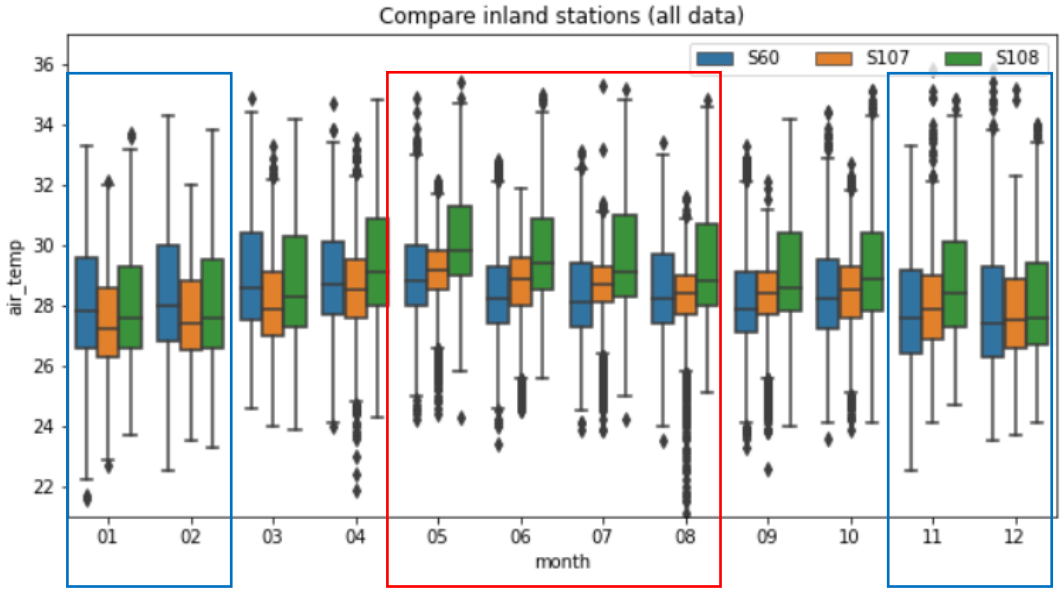




Define Scope of Data For Analysis

- Define season group – hot months vs cool months.
- Sitting on the equator, there is no season here except that the period from Nov to Feb are cooler.
- The study will be split between hot month group (May to Aug) and cool month group (Nov to Feb).



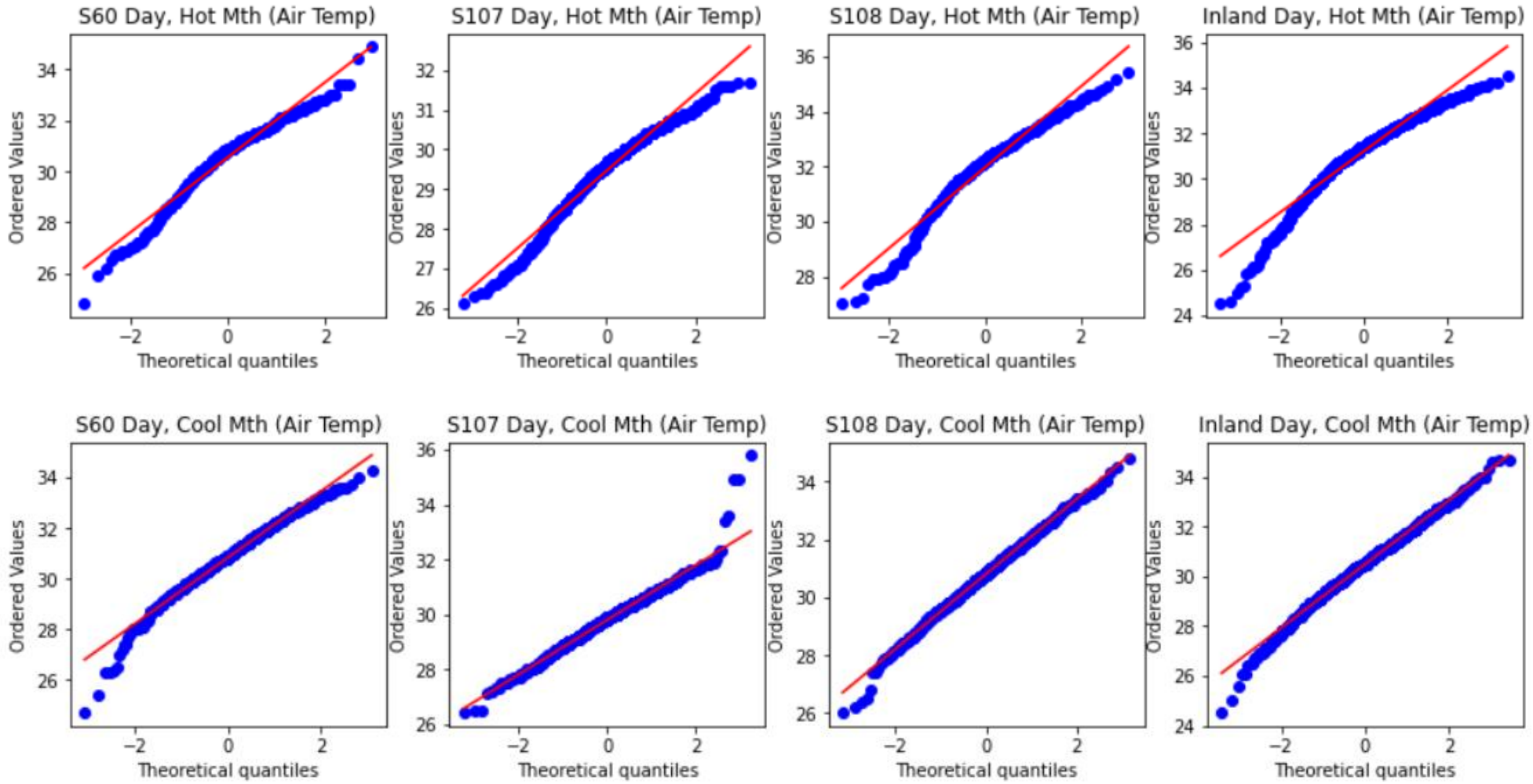


Data Analysis

- Test 1 (Day): to compare if each of the identified weather station near the coast has the same temperature as inland based on the following conditions:
  - The same time period of 12 – 3pm
  - High UV index of 6 and above
  - No rain (to remove noise from the data due to rain)
- Mood's Median Test – distributions are not normally distributed:
  - H0: $station@coast_{med\_temp} = inland_{med\_temp}$
  - H1: $station@coast_{med\_temp} \neq inland_{med\_temp}$
- Results:
  - The tests show significant difference in median temperature between all 3 stations and inland at the hottest time of the day at high UV index level (no rain).
  - S107 (East Coast Parkway) is coolest among all at the hottest time of the day (median).
  - Highest median temperature is at the coastal area for both month groups.
  - Coastal areas are more humid than inland in the period from 12 – 3 pm.
  - Wind speed could be higher at coast at times but the median is similar to inland. Hence wind is not sufficient to mitigate the higher humidity on the whole.

Data Analysis

- Check normality of the distribution – all not normally distributed



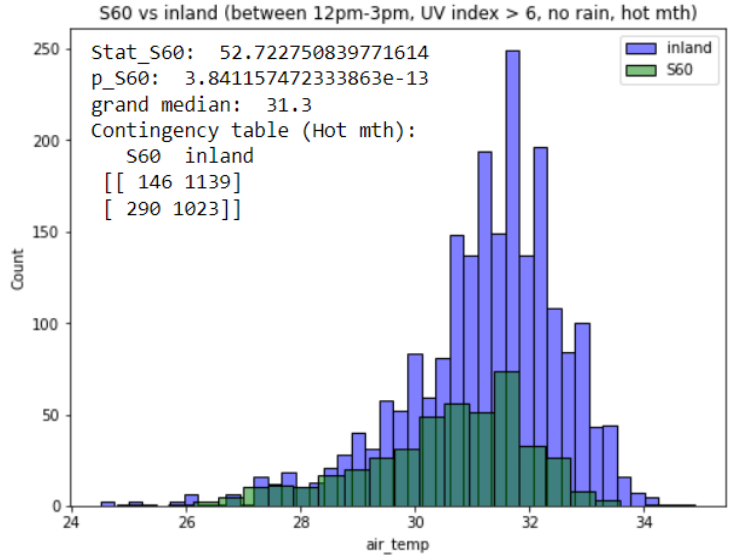
Data Analysis

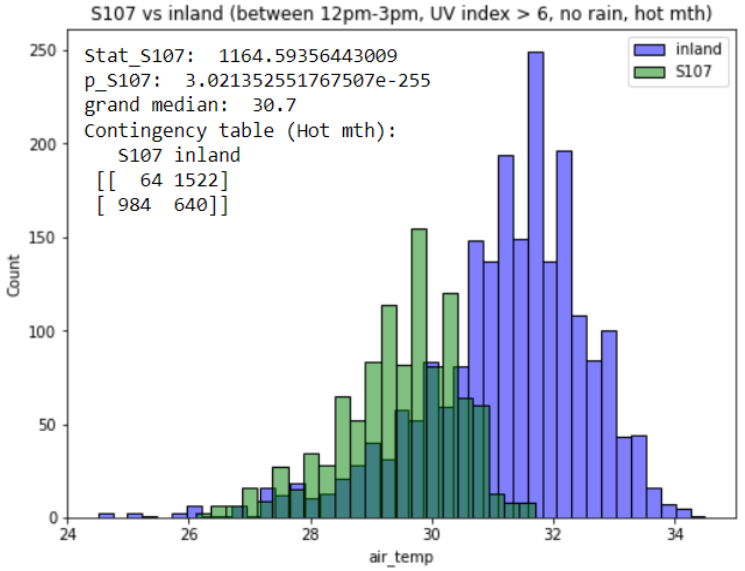
S60: Sentosa
S107: East Coast Parkway
S108: Marina Gardens Drive

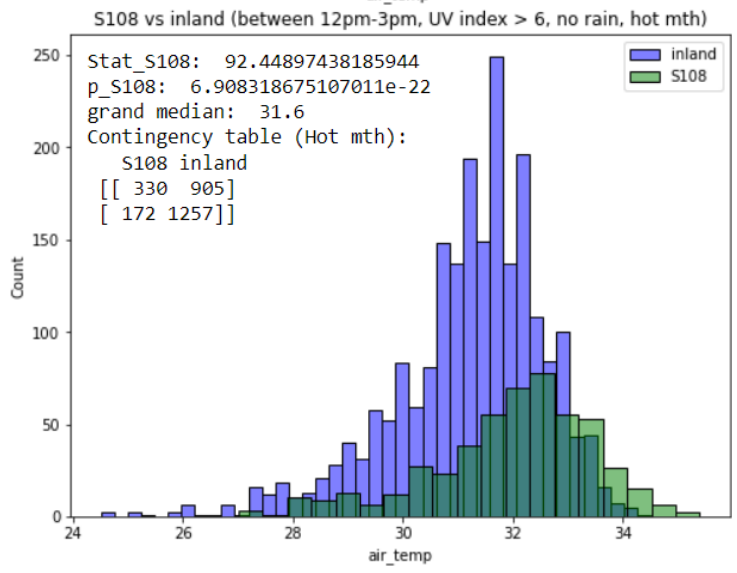
Day, Hot month period:

- Comparison of temperature shows each of the stations near the coast is significantly different from inland.
- Not all stations at the coast are cooler than inland.
- Lowest median temp: S107
- Highest median temp: S108
- Lowest at 10th-pct: S107
- Highest at 90th-pct: S108

location	count	mean	std	min	10%	25%	50%	75%	90%	max
Inland	2162.0	31.21	1.38	24.5	29.40	30.50	31.4	32.1	32.7	34.5
S107	1048.0	29.45	0.99	26.1	28.10	28.90	29.6	30.2	30.6	31.7
S108	502.0	31.97	1.50	27.0	29.91	31.22	32.2	33.0	33.6	35.4
S60	436.0	30.54	1.49	24.8	28.50	29.78	30.8	31.5	32.2	34.9







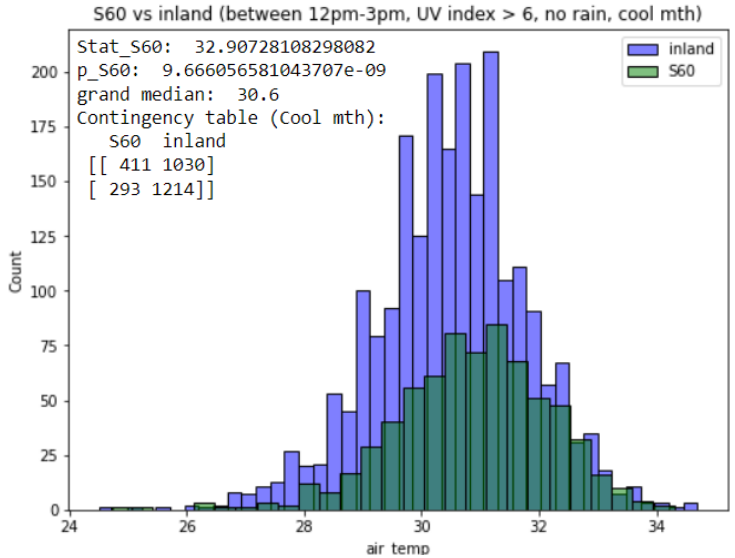
Data Analysis

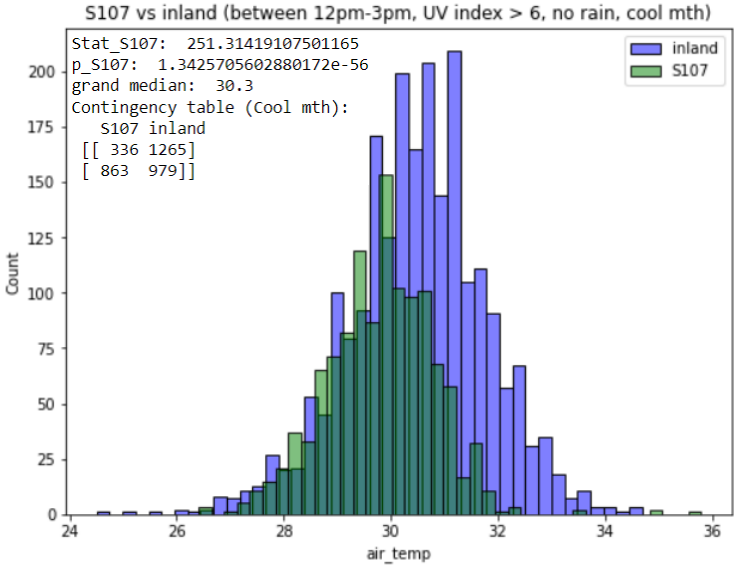
S60: Sentosa
S107: East Coast Parkway
S108: Marina Gardens Drive

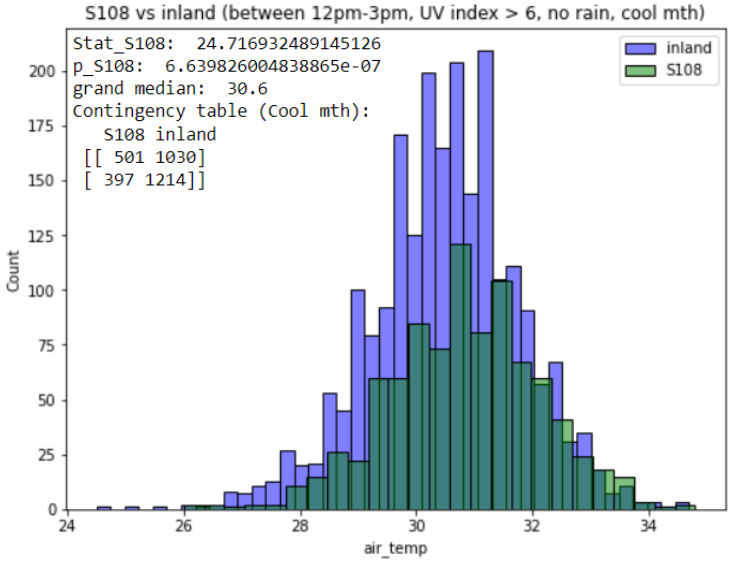
Day, Cool month period:

- Comparison of temperature shows each of the stations near the coast is significantly different from inland.
- Not all stations at the coast are cooler than inland.
- Lowest median temp: S107
- Highest median temp: S60
- Lowest at 10th-pct: S107
- Highest at 90th-pct: S60

location	count	mean	std	min	10%	25%	50%	75%	90%	max
Inland	2244.0	30.49	1.28	24.5	28.90	29.7	30.5	31.3	32.1	34.7
S107	1199.0	29.80	1.00	26.4	28.50	29.1	29.9	30.4	31.0	35.8
S108	898.0	30.81	1.29	26.0	29.20	30.0	30.8	31.7	32.4	34.8
S60	704.0	30.85	1.31	24.7	29.23	30.1	30.9	31.7	32.5	34.3



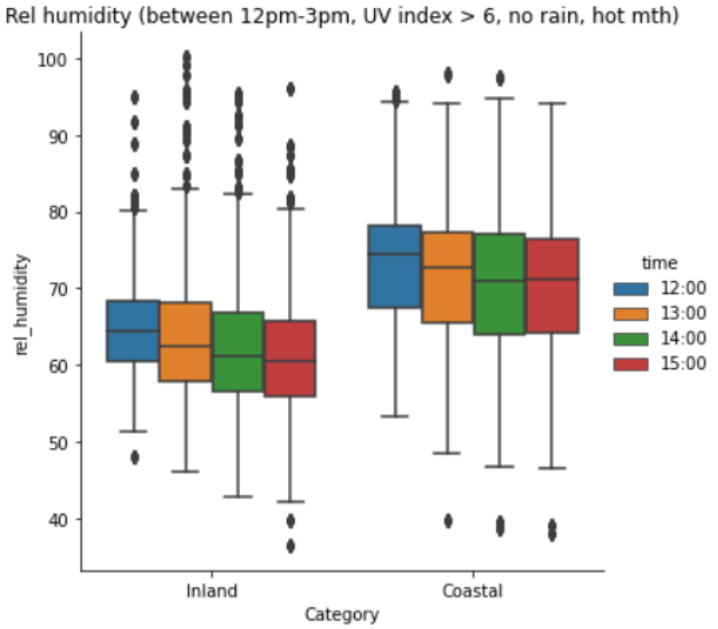


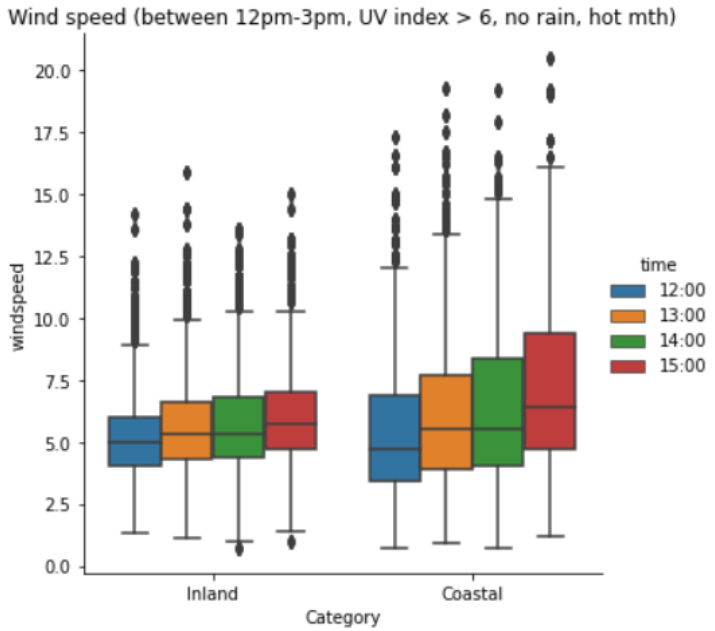


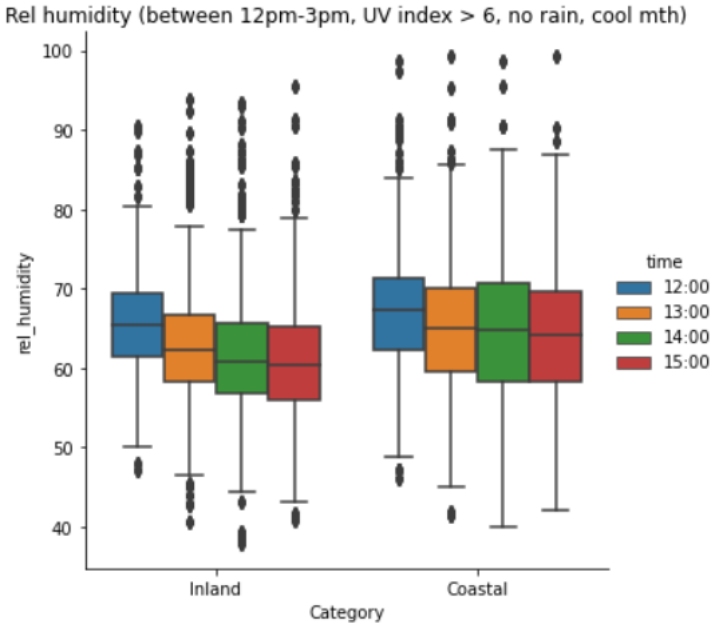
Data Analysis

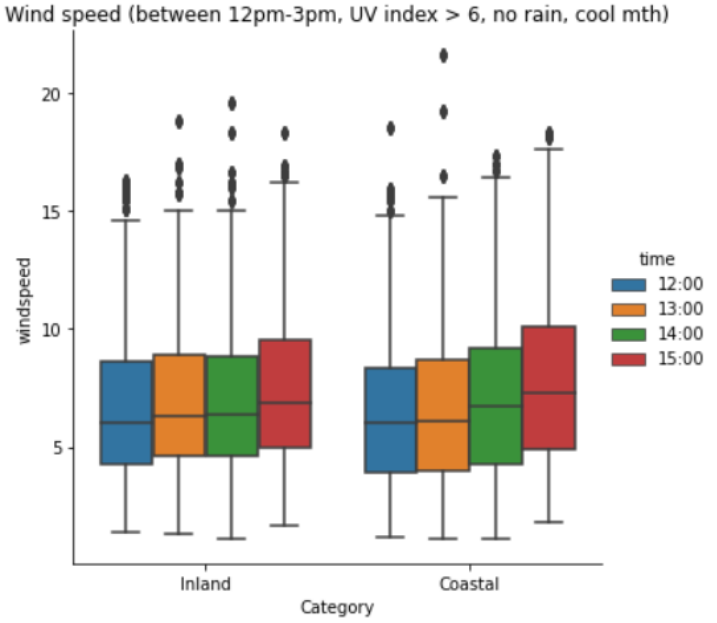
Between 12 - 3pm:

- On the whole, relative humidity is higher at the coastal areas especially during hot month group (May-Aug).
- A higher relative humidity value means one would feel stickier at the same temperature.
- Coastal areas has higher occurrences of stronger wind for the hot month group but the median remains comparable to inland.







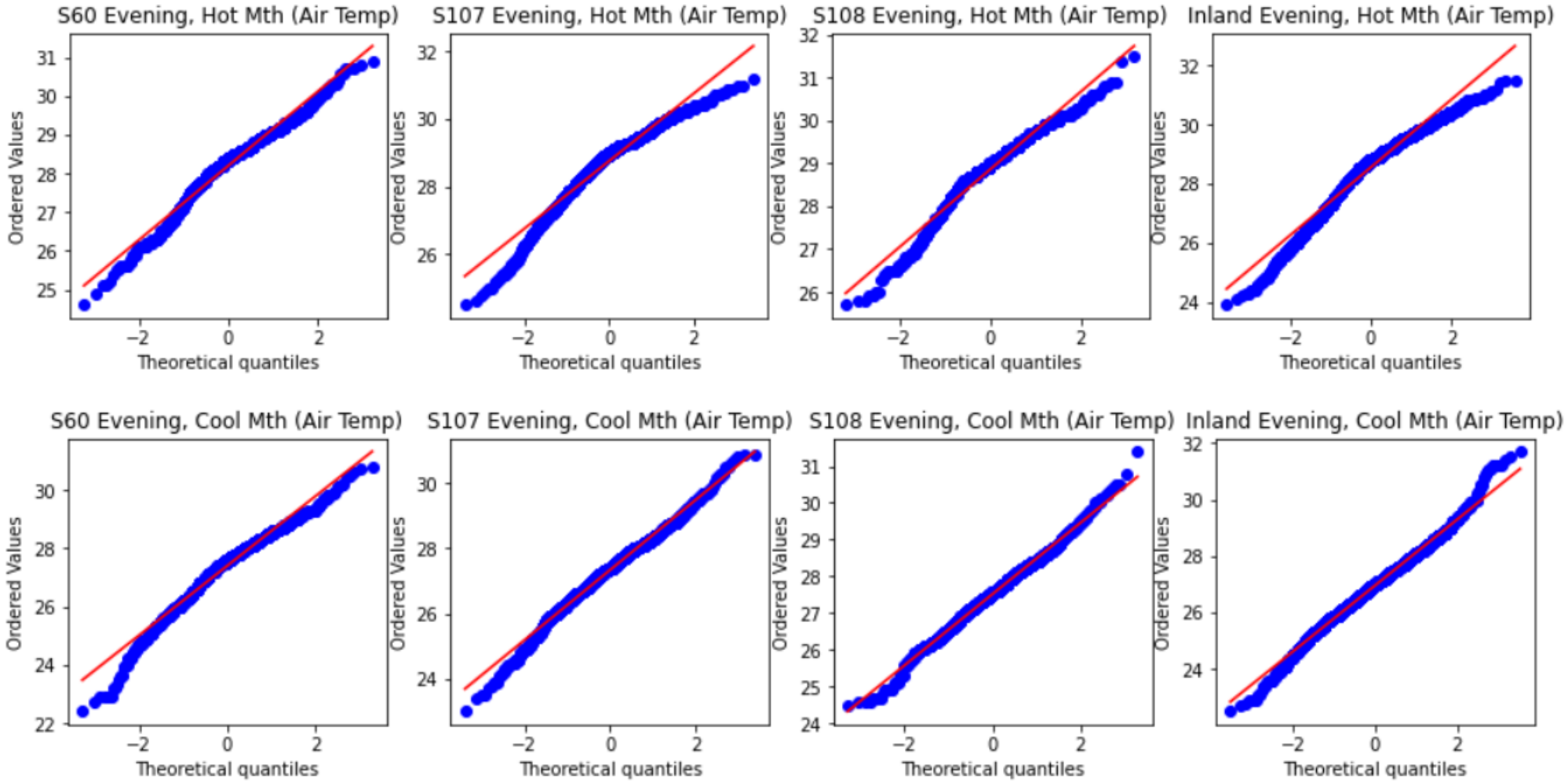


Data Analysis

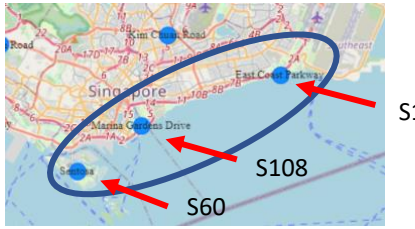
- Test 2 (Evening): to compare if each of the identified weather station near the coast has the same temperature as inland based on the following conditions:
  - The same time period of 8pm – 11pm
  - No rain
- Mood's Median Test – distributions are not normally distributed:
  - H0: $station@coast_{med\_temp} = inland_{med\_temp}$
  - H1: $station@coast_{med\_temp} \neq inland_{med\_temp}$
- Results:
  - The tests show significant difference in median temperature between all 3 stations and inland in the night.
  - Inland is coolest during the cooler months (median).
  - Highest median temperature is at the coastal area for both month groups.
  - Wind speed is higher at coast at times but the median is similar to inland.

Data Analysis

- Check normality of the distribution – all not normally distributed



Data Analysis

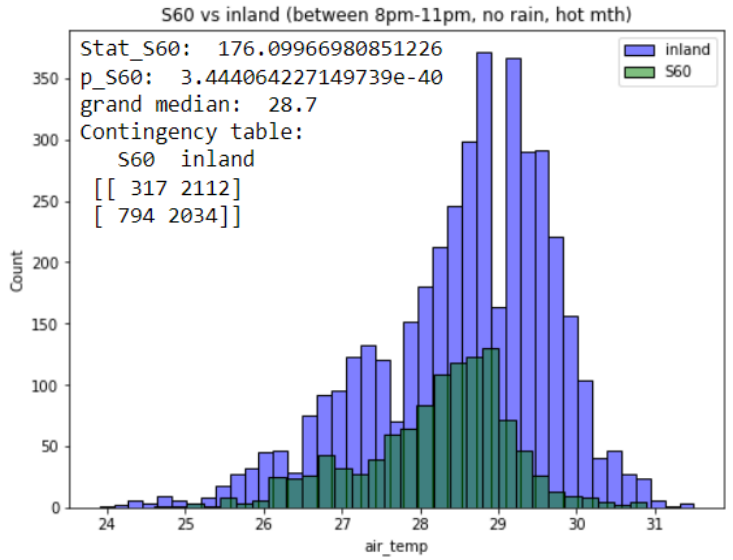


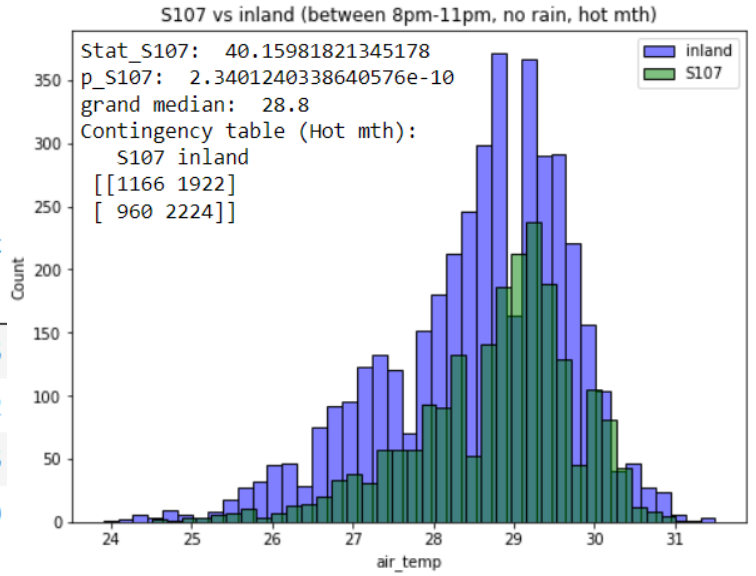
S60: Sentosa
S107: East Coast Parkway
S108: Marina Gardens Drive

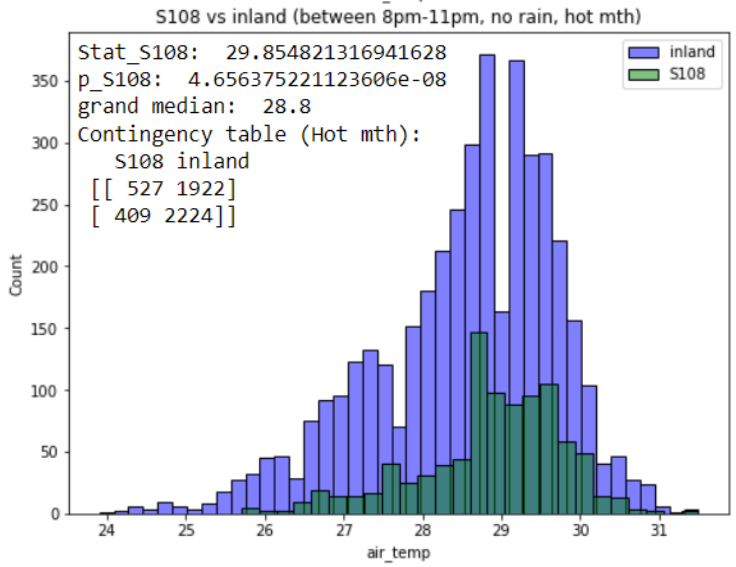
Evening, Hot month period:

- Comparison of temperature shows each of the stations near the coast is significantly different from inland.
- Not all stations at the coast are cooler than inland.
- Lowest median temp: S60
- Highest median temp: S107 / S108
- Lowest at 10th-pct: S60
- Highest at 90th-pct: S107 / S108

location	count	mean	std	min	10%	25%	50%	75%	90%	max
Inland	4146.0	28.55	1.16	23.9	26.9	27.9	28.8	29.4	29.8	31.5
S107	2126.0	28.76	1.02	24.5	27.3	28.2	29.0	29.4	29.9	31.2
S108	936.0	28.86	0.92	25.7	27.5	28.4	29.0	29.5	29.9	31.5
S60	1111.0	28.20	0.97	24.6	26.7	27.7	28.4	28.8	29.3	30.9







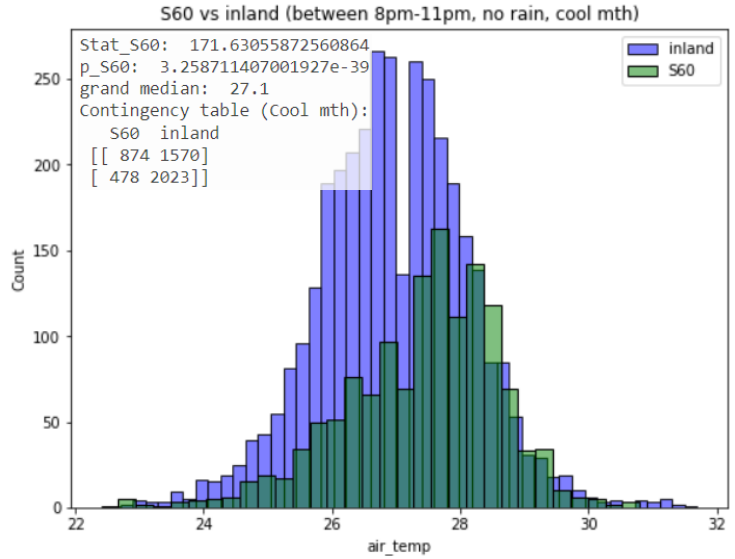
Data Analysis

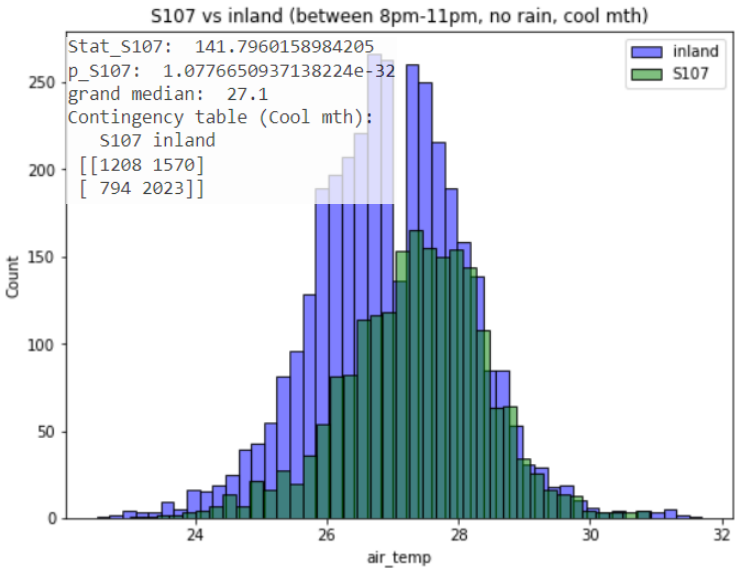
S60: Sentosa
S107: East Coast Parkway
S108: Marina Gardens Drive

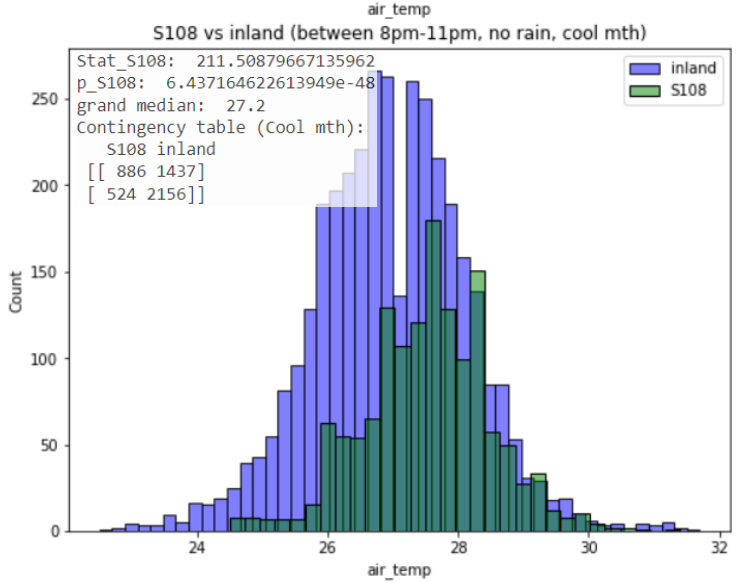
Evening, Cool month period:

- Comparison of temperature shows each of the stations near the coast is significantly different from inland.
- Not all stations at the coast are cooler than inland.
- Lowest median temp: Inland
- Highest median temp: S60 / S108
- Lowest at 10th-pct: Inland
- Highest at 90th-pct: S60 / S108

location	count	mean	std	min	10%	25%	50%	75%	90%	max
Inland	3593.0	26.96	1.16	22.5	25.5	26.2	27.0	27.7	28.3	31.7
S107	2002.0	27.34	1.08	23.0	26.0	26.7	27.4	28.0	28.6	30.9
S108	1410.0	27.52	0.97	24.5	26.2	26.9	27.6	28.2	28.7	31.4
S60	1352.0	27.40	1.21	22.4	25.8	26.7	27.6	28.2	28.7	30.8



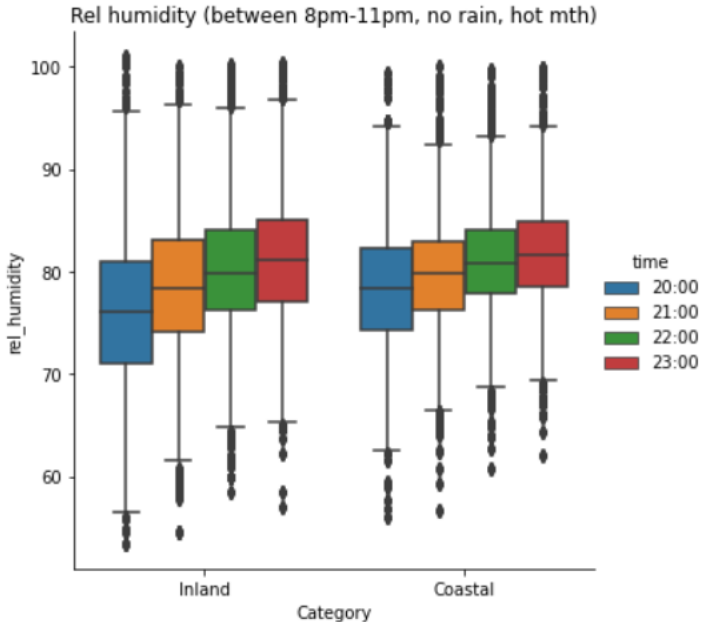


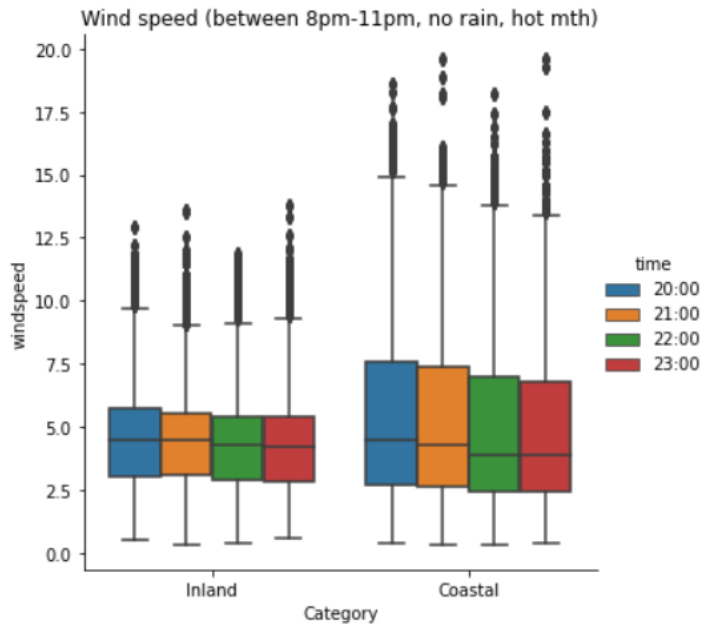


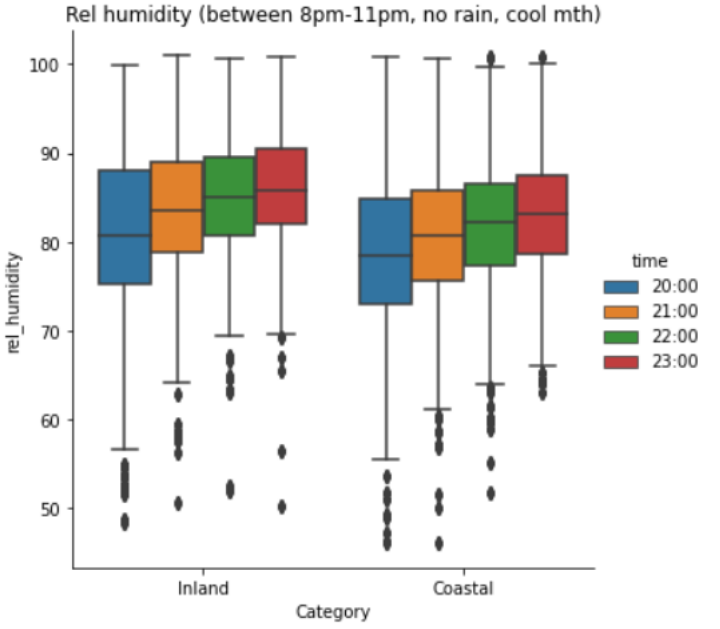
Data Analysis

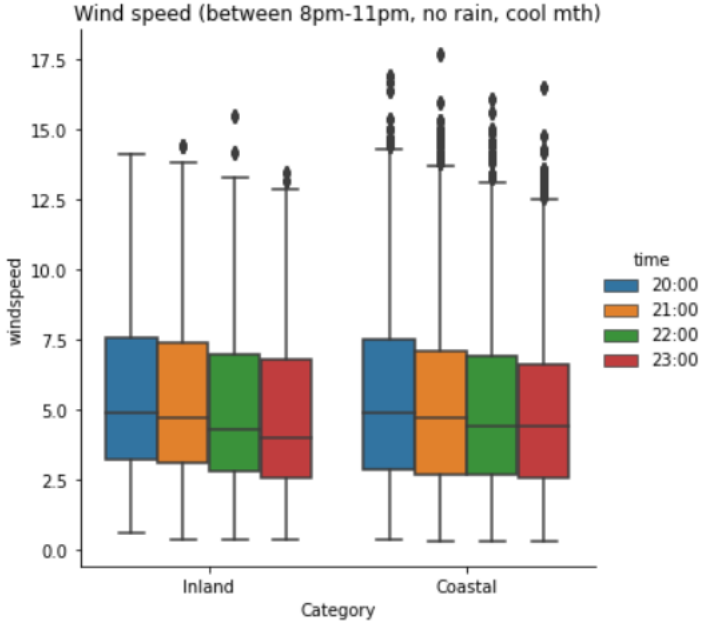
Between 8 - 11pm:

- Relative humidity is comparable between inland and coastal for both hot and cool month groups.
- Coastal areas has higher occurrences of stronger wind during the hotter months but the median remains comparable to inland.







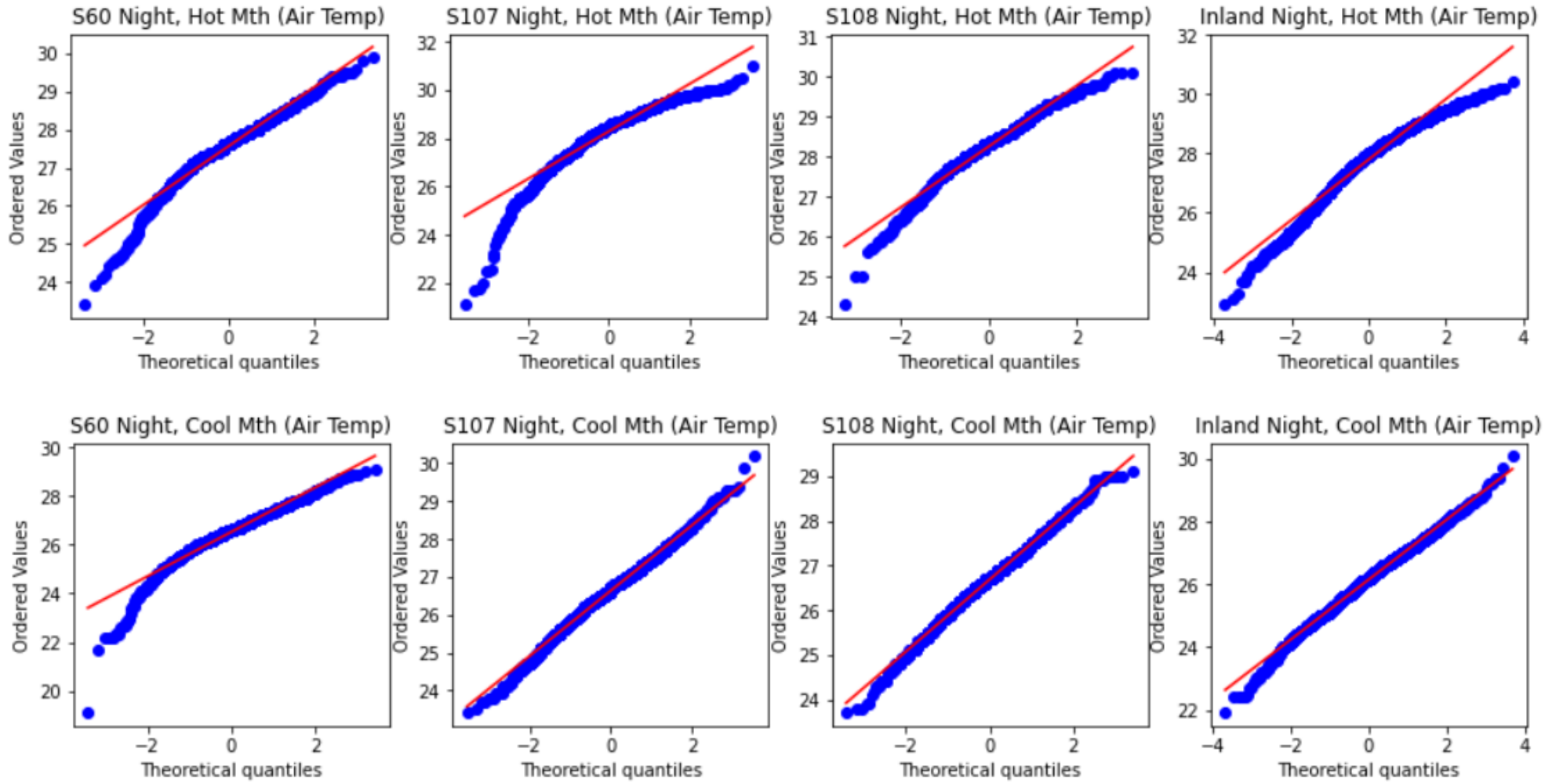


Data Analysis

- Test 3 (Night): to compare if each of the identified weather station near the coast has the same temperature as inland based on the following conditions:
  - The same time period of 12am – 6am
  - No rain
- Mood's Median Test – distributions are not normally distributed:
  - H0: $station@coast_{med\_temp} = inland_{med\_temp}$
  - H1: $station@coast_{med\_temp} \neq inland_{med\_temp}$
- Results:
  - The tests show significant difference in median temperature between all 3 stations and inland at past midnight.
  - Inland is coolest during the cooler months (median).
  - Highest median temperature is at the coastal area for both month groups.
  - Relative humidity is comparable between coastal and inland for both month groups.
  - Wind speed is slightly higher at coast at times but the median is similar to inland.

Data Analysis

- Check normality of the distribution – all not normally distributed

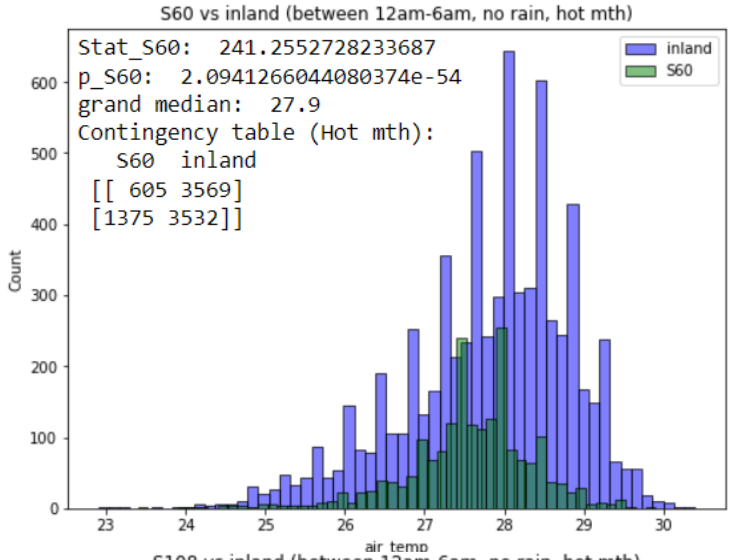


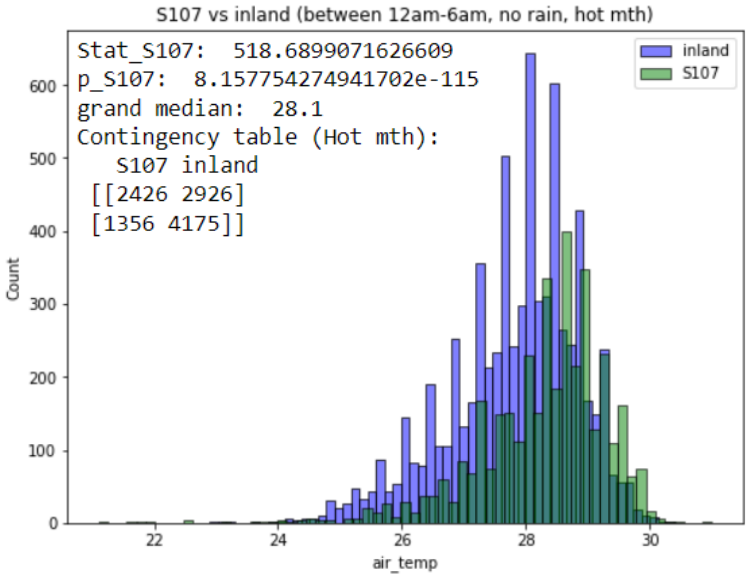
Data Analysis

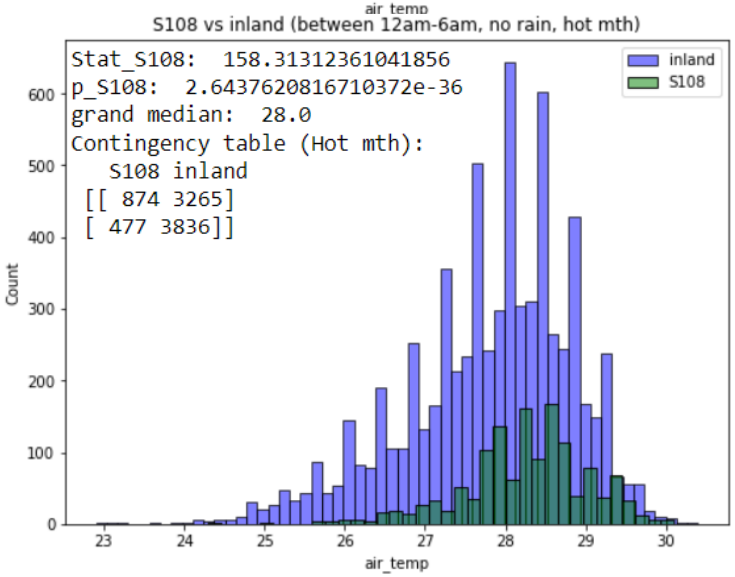
Night, Hot month period:

- Comparison of temperature shows each of the stations near the coast is significantly different from inland.
- Not all stations at the coast are cooler than inland.
- Lowest median temp: S60
- Lowest at 10th-pct: Inland
- Highest at 90th-pct: S107

location	count	mean	std	min	10%	25%	50%	75%	90%	max
Inland	7101.0	27.79	1.04	22.9	26.3	27.2	28.0	28.5	29.0	30.4
S107	3782.0	28.28	1.02	21.1	27.0	27.8	28.5	29.0	29.4	31.0
S108	1351.0	28.26	0.77	24.3	27.3	27.8	28.3	28.7	29.2	30.1
S60	1980.0	27.57	0.78	23.4	26.6	27.2	27.6	28.0	28.5	29.9





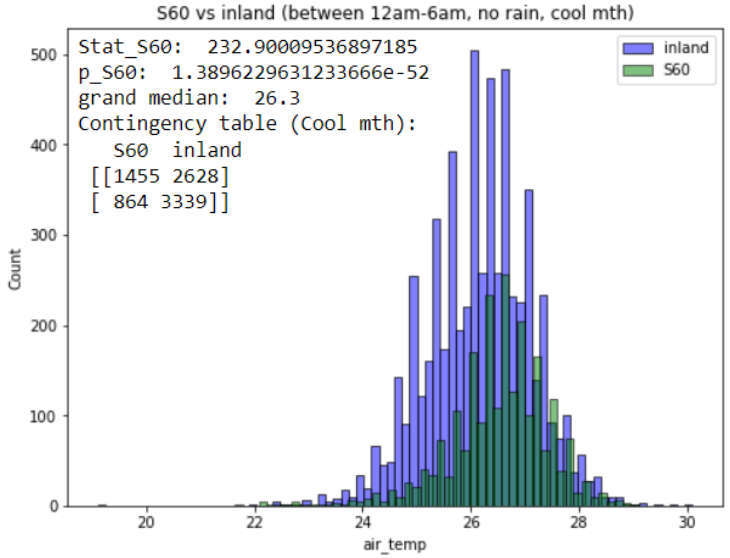


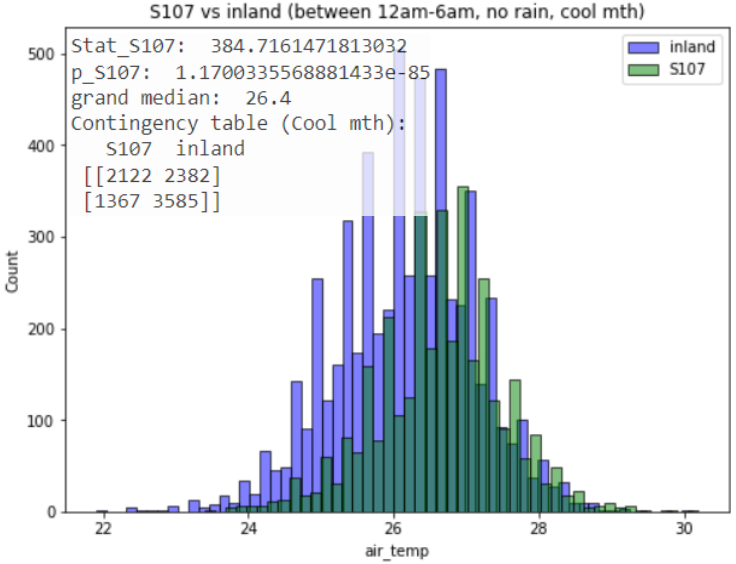
Data Analysis

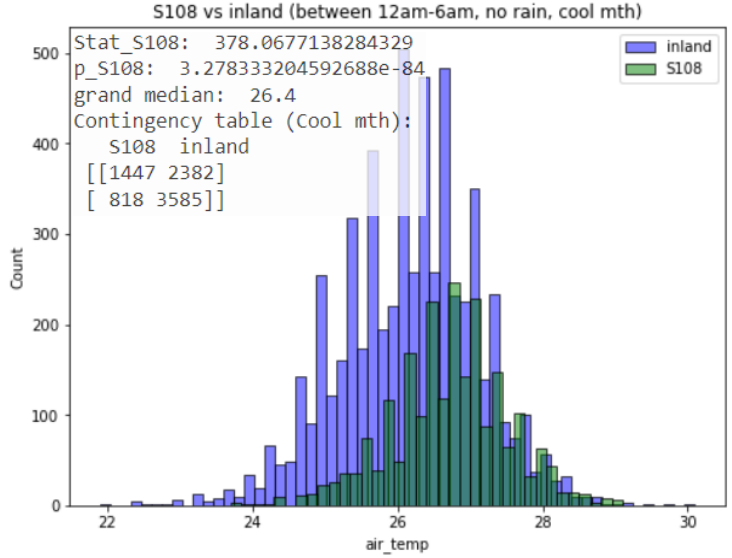
Night, Cool month period:

- Comparison of temperature shows each of the stations near the coast is significantly different from inland.
- Stations at the coast are warmer than inland.
- Lowest median temp: Inland
- Lowest at 10th-pct: S60
- Highest at 90th-pct: S107 / S108

location	count	mean	std	min	10%	25%	50%	75%	90%	max
Inland	5967.0	26.16	0.96	21.9	24.9	25.5	26.2	26.8	27.3	30.1
S107	3489.0	26.63	0.87	23.4	25.5	26.1	26.7	27.2	27.7	30.2
S108	2265.0	26.69	0.81	23.7	25.6	26.2	26.7	27.2	27.7	29.1
S60	2319.0	26.53	0.93	19.1	25.4	26.1	26.6	27.1	27.6	29.1

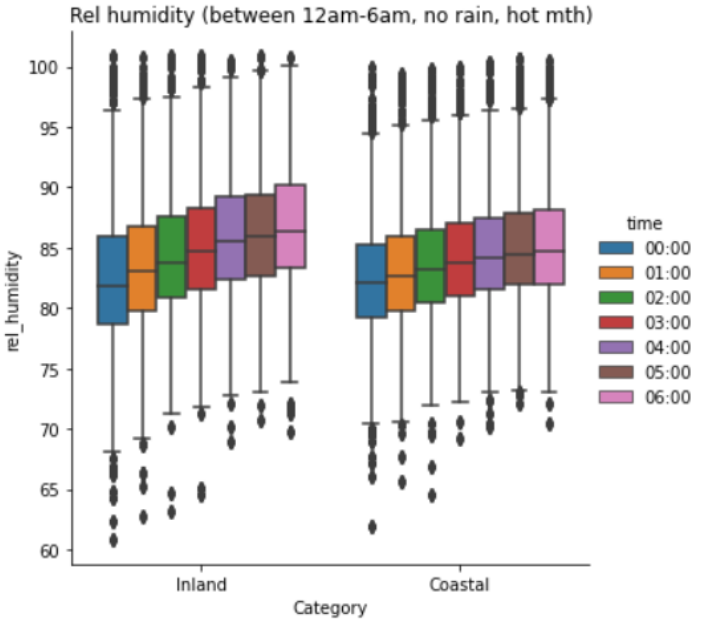


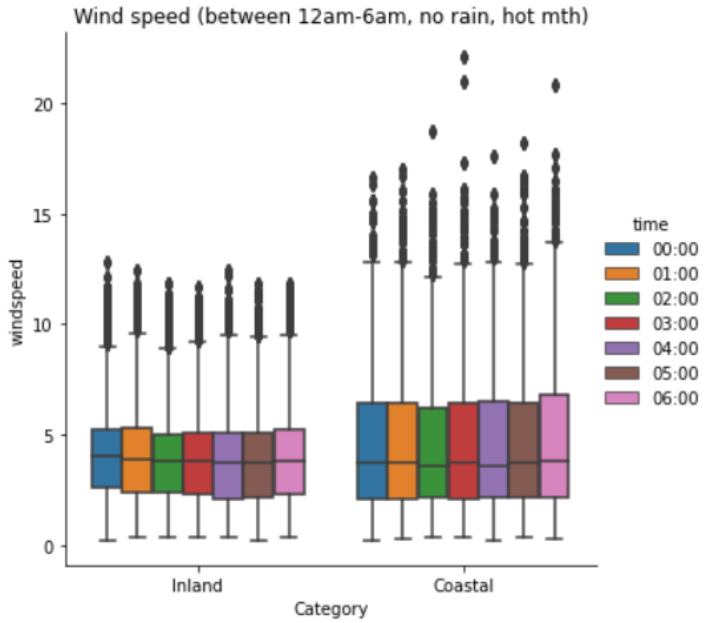


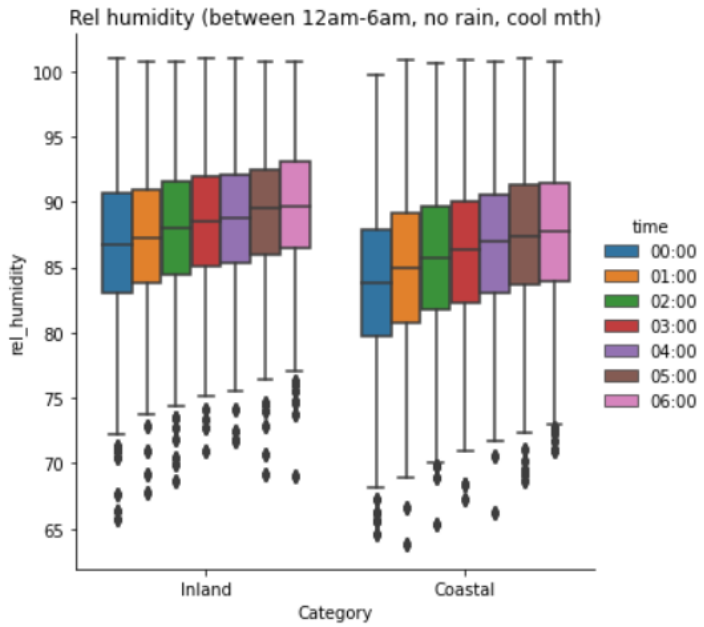


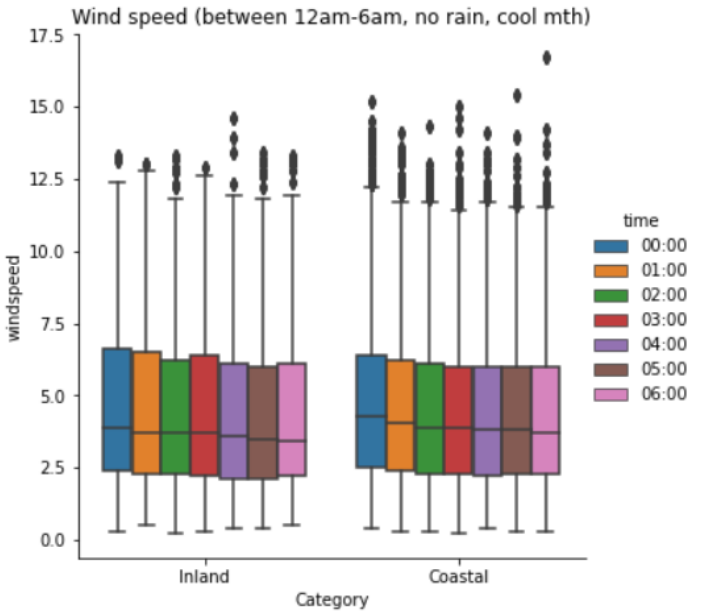
Data Analysis

- Relative humidity is comparable between inland and coastal.
- Coastal areas has slight occurrences of stronger wind but the median remains comparable to inland.









Conclusion

There is indeed significant difference in temperatures between coastal and inland areas however the data refuted the expectation that the coastal areas will certainly be cooler than inland.

Stations along the coast do not experience similar weather conditions even though all 3 line up along the southeast of the island.

The largest difference between the stations is 2.6˚C for 12-3pm/hot month group, coolest at S107 and warmest at S108.

Humidity is higher at the coast during the hottest time of day during the warmer months; that means to expect some stickiness in the salty air during that time.

To summarise the test outcome:

		Temperature (median)	Rel humidity	Wind Speed
12-3pm	Hot mth	coast not always cooler	more humid	at times higher. Same median
	Cool Mth	coast not always cooler	slightly more humid	comparable
8-11pm	Hot mth	coast not always cooler	slightly more humid	at times higher. Same median
	Cool Mth	inland is coolest	comparable	comparable
12-6am	Hot mth	coast not always cooler	comparable	at times higher. Same median
	Cool Mth	inland is coolest	slightly less humid	comparable

- *So, moving to the coast does not necessarily mean enjoying cool breeze but yet could feel more sticky. If I were to insist, which is the best among these few locations?*

Conclusion

To help with the decision making, temperature is ranked and tallied, putting humidity aside:

			S60	S107	S108	Inland
Day	Hot mth	10th	2	1	4	3
12-3pm		median	2	1	4	3
		90th	2	1	4	3
	Cool mth	10th	4	1	3	2
		median	4	1	3	2
		90th	4	1	3	2
Night	Hot mth	10th	1	3	4	2
8-11pm		median	1	3	3	2
		90th	1	3	3	2
	Cool mth	10th	2	3	4	1
		median	3	2	3	1
		90th	3	2	3	1
Midnight	Hot mth	10th	2	3	4	1
12-6am		median	1	4	3	2
		90th	1	4	3	2
	Cool mth	10th	2	3	4	1
		median	2	3	3	1
		90th	2	3	3	1
	Total		**39**	**42**	**61**	**32**

- Inland is the coolest!
- S108 (Marina Gardens Drive) is almost always the warmest here.

So I'm keeping my cheque!

Recommendation

Other studies that can be done with this dataset:

1) Pursue similar comparison with other parts of Singapore.
   - Against other towns
   - Against areas that are near to nature reserves
   - Island against island: Pulau Ubin vs Sentosa vs Singapore

2) Study if the heavy industries areas experience higher temperature.

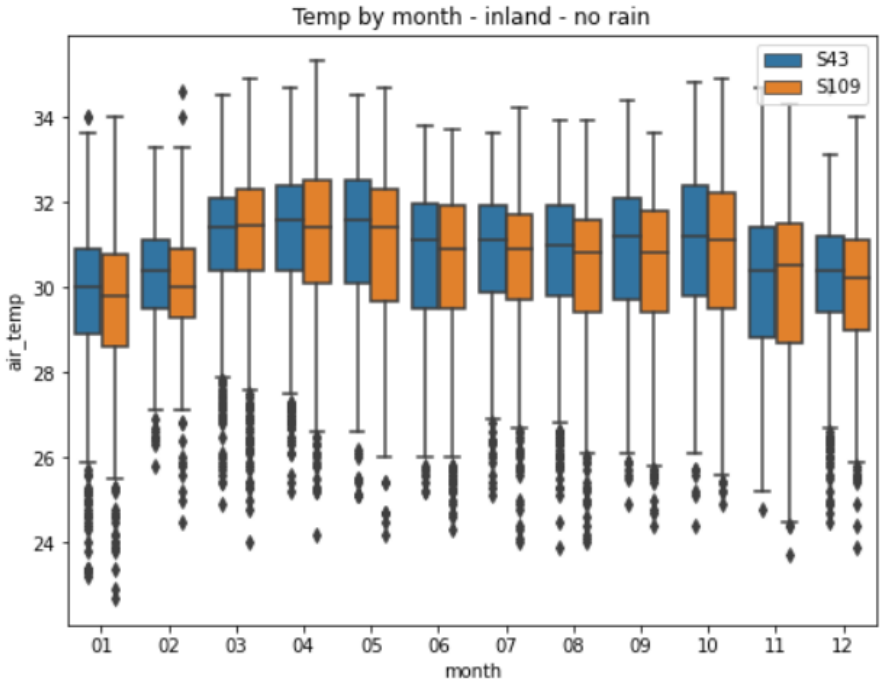
3) Combine this dataset with other datasets property price and population density
   -> *this might help me find my ideal new home with the right price and the least bustling!*

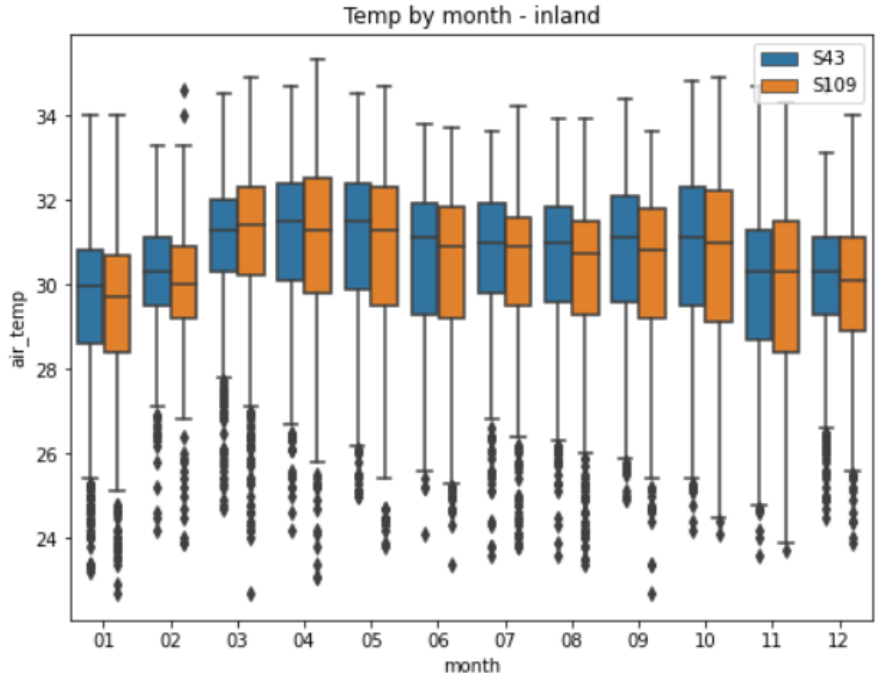
Acknowledgment

1) Data source: https://data.gov.sg/developer
2) Extreme temperatures in Singapore history: http://www.weather.gov.sg/climate-historical-extremes-temperature
3) UV Index info: https://www.nea.gov.sg/weather/ultraviolet-index/uv-radiation-uv-index
4) Metrological Service Singapore. Information on the weather stations: http://www.weather.gov.sg/learn_observations/
5) Cover photo courtesy of Jovyn Lim (Central Singapore, Jan 2022)

Additional Info

- Inland data without rain vs with rain





Additional Info

- Coastal data without rain vs with rain