Looked through local data - not much weather stations available for comparison...

TORONTO CITY
ONTARIO
Current Station Operator: ECCC - MSC

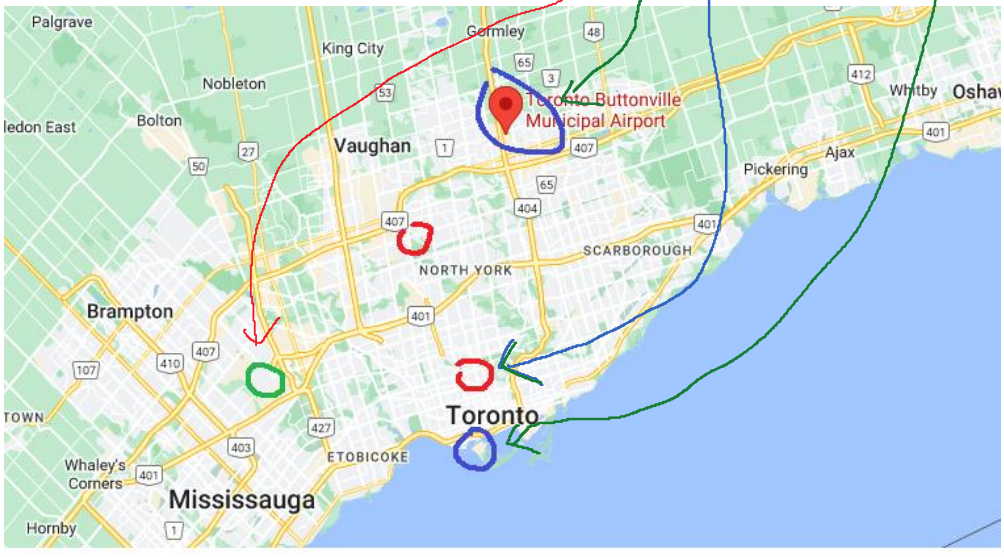
TORONTO CITY CENTRE
ONTARIO
ırrent Station Operator: NAVCAN

TORONTO INTL A
ONTARIO
Current Station Operator: NAVCAN

TORONTO NORTH YORK
ONTARIO
Current Station Operator: CCN

TORONTO BUTTONVILLE A
ONTARIO
Current Station Operator: NAVCAN

Longitude: 79°22'07.000" W Ele



Monday, May 16, 2022 5:12 PM

nasa - has solar parameters
- nrcan

May 30 meeting

Monday, May 30, 2022 5:00 PM

find out what dad does for thesis

retscreen key

get solar radiation level from retscreen
weather label

what exact models do you want me to model?

- ANN
- LSTM

Should I try from just using the temperature,
or should i factor in dew point temp, pressure

- Temperature, humidex, wind chill, weather label, rel hum, dew point

retscan → canmedio?

crossvalidation?

hyperparameter tuning?

Are there any references to compare my MAE/MAPE results to

time period of training? jan 2016 - dec 31 2021 data

- create lags

outliers - remove

retscan -> needs license, but it only gives weather data for certain stations

eda

Sunday, June 5, 2022 3:53 PM

windchill → just temp x wind speed when temperature is below 0

- since I have both features, I don't think adding correlated features will help the model

humidex → air temp with dew point

- since I have both features, I don’t think adding this correlated feature will help

added lag → not sure how much lag, so I addeed 3 hours of lagged data for now

notes

Monday, June 6, 2022 5:21 PM

try adding in humidex + wind chill
- wind chill make it equal to dry box temperature if cant leave as empty

add wind direction

- add solar radiation

- check correlations

check other weather modelling for lit review

future:
- add in day time / night time parameter
- or maybe winter/summer dividing of model
- see if more data / less data improves model

Monday, June 20, 2022 5:10 PM

try to add humidex - done
solar radiation - done
split data in winter and summer
day time, night time

SHAP values
tscv

add LAT/LONG
see if you can get data from retscreen anywhere - done

neural networks

add lag to humidex, solar_rad done

compare the data I get vs the station data to see if the data is different done

do feature engineering like PCA to see which features are good

tscv
hpt

lat long → not added
retscreen data → varies based on location but I'm not sure by what criteria it varies

Canada - ON - Toronto (43.7°N, -79.4°E) Elevation: 108 m (Facility)

Period	Begin	End	Duration Hours	Air temperature - average °C	Relative humidity %	Precipitation mm	Solar radiation - horizontal Wh/m²	Atmospheric pressure kPa	Wind speed m/s	Earth temperature °C	Comments
1	2012-12-30 19:00	2012-12-30 20:00	1	-0.9	83.2%	0.2	0	101.0	7.4	2.1	
2	2012-12-30 20:00	2012-12-30 21:00	1	-1.3	84.4%	0.1	0	101.0	7.6	1.7	
3	2012-12-30 21:00	2012-12-30 22:00	1	-1.3	85.3%	0.1	0	101.0	8.2	1.8	
4	2012-12-30 22:00	2012-12-30 23:00	1	-1.3	86.1%	0.1	0	101.0	8.5	2.0	
5	2012-12-30 23:00	2012-12-31 00:00	1	-1.5	87.6%	0.1	0	100.9	8.8	2.1	
6	2012-12-31 00:00	2012-12-31 01:00	1	-1.8	89.4%	0.0	0	100.8	8.8	1.9	
7	2012-12-31 01:00	2012-12-31 02:00	1	-2.0	90.3%	0.0	0	100.8	9.1	1.9	
8	2012-12-31 02:00	2012-12-31 03:00	1	-1.9	90.5%	0.0	0	100.7	9.0	1.9	
9	2012-12-31 03:00	2012-12-31 04:00	1	-1.7	90.9%	0.0	0	100.6	9.0	1.9	
10	2012-12-31 04:00	2012-12-31 05:00	1	-1.3	91.1%	0.0	0	100.6	8.9	2.1	
11	2012-12-31 05:00	2012-12-31 06:00	1	-0.6	90.9%	0.0	0	100.5	8.8	2.3	
12	2012-12-31 06:00	2012-12-31 07:00	1	0.0	90.0%	0.0	0	100.4	8.9	2.5	
13	2012-12-31 07:00	2012-12-31 08:00	1	0.4	89.5%	0.0	0	100.4	9.1	2.6	
14	2012-12-31 08:00	2012-12-31 09:00	1	0.8	88.6%	0.0	37	100.3	9.5	2.8	
15	2012-12-31 09:00	2012-12-31 10:00	1	1.3	86.6%	0.0	125	100.2	10.4	3.2	
16	2012-12-31 10:00	2012-12-31 11:00	1	1.7	83.3%	0.0	206	100.2	11.1	3.6	

Canada - ON - Toronto (43.8°N, -79.4°E) Elevation: 166 m (Facility)

Period	Begin	End	Duration Hours	Air temperature - average °C	Relative humidity %	Precipitation mm	Solar radiation - horizontal Wh/m²	Atmospheric pressure kPa	Wind speed m/s	Earth temperature °C	Comments
1	2012-12-30 19:00	2012-12-30 20:00	1	-5.5	94.7%	0.1	0	100.2	4.6	-5.9	
2	2012-12-30 20:00	2012-12-30 21:00	1	-6.5	97.6%	0.0	0	100.2	5.0	-6.9	
3	2012-12-30 21:00	2012-12-30 22:00	1	-7.1	99.3%	0.0	0	100.2	5.4	-7.4	
4	2012-12-30 22:00	2012-12-30 23:00	1	-7.2	99.6%	0.0	0	100.2	5.4	-7.4	
5	2012-12-30 23:00	2012-12-31 00:00	1	-7.0	98.6%	0.0	0	100.1	5.3	-6.9	
6	2012-12-31 00:00	2012-12-31 01:00	1	-6.6	97.6%	0.0	0	100.0	5.2	-6.3	
7	2012-12-31 01:00	2012-12-31 02:00	1	-6.3	96.8%	0.0	0	99.9	5.4	-6.0	
8	2012-12-31 02:00	2012-12-31 03:00	1	-6.1	96.1%	0.0	0	99.9	5.3	-5.8	
9	2012-12-31 03:00	2012-12-31 04:00	1	-6.2	96.1%	0.0	0	99.8	5.4	-5.9	
10	2012-12-31 04:00	2012-12-31 05:00	1	-6.1	96.3%	0.0	0	99.7	5.2	-5.8	
11	2012-12-31 05:00	2012-12-31 06:00	1	-5.7	96.4%	0.0	0	99.7	5.0	-5.3	
12	2012-12-31 06:00	2012-12-31 07:00	1	-5.3	96.4%	0.0	0	99.6	5.2	-4.9	
13	2012-12-31 07:00	2012-12-31 08:00	1	-5.2	96.9%	0.0	0	99.6	5.1	-4.9	
14	2012-12-31 08:00	2012-12-31 09:00	1	-4.8	96.4%	0.0	37	99.5	5.5	-4.4	
15	2012-12-31 09:00	2012-12-31 10:00	1	-3.6	93.6%	0.0	125	99.4	6.8	-2.9	
16	2012-12-31 10:00	2012-12-31 11:00	1	-2.4	89.6%	0.0	206	99.3	7.8	-1.5	
17	2012-12-31 11:00	2012-12-31 12:00	1	-1.4	86.0%	0.0	228	99.2	8.0	-0.5	
18	2012-12-31 12:00	2012-12-31 13:00	1	0.8	84.0%	0.0	188	99.1	8.0	0.2	

Canada - ON - Toronto (43.7°N, -79.3°E) Elevation: 100 m (Facility)

Period	Begin	End	Duration Hours	Air temperature - average °C	Relative humidity %	Precipitation mm	Solar radiation - horizontal Wh/m²	Atmospheric pressure kPa	Wind speed m/s	Earth temperature °C	Co
1	2012-12-30 19:00	2012-12-30 20:00	1	-0.9	83.2%	0.2	0	101.1	7.4	2.1	
2	2012-12-30 20:00	2012-12-30 21:00	1	-1.3	84.4%	0.1	0	101.1	7.6	1.7	
3	2012-12-30 21:00	2012-12-30 22:00	1	-1.3	85.3%	0.1	0	101.1	8.2	1.8	
4	2012-12-30 22:00	2012-12-30 23:00	1	-1.3	86.1%	0.1	0	101.1	8.5	2.0	
5	2012-12-30 23:00	2012-12-31 00:00	1	-1.5	87.6%	0.1	0	101.0	8.8	2.1	
6	2012-12-31 00:00	2012-12-31 01:00	1	-1.8	89.4%	0.0	0	100.9	8.8	1.9	
7	2012-12-31 01:00	2012-12-31 02:00	1	-2.0	90.3%	0.0	0	100.9	9.1	1.9	
8	2012-12-31 02:00	2012-12-31 03:00	1	-1.9	90.5%	0.0	0	100.8	9.0	1.9	
9	2012-12-31 03:00	2012-12-31 04:00	1	-1.7	90.9%	0.0	0	100.7	9.0	1.9	
10	2012-12-31 04:00	2012-12-31 05:00	1	-1.3	91.1%	0.0	0	100.7	8.9	2.1	
11	2012-12-31 05:00	2012-12-31 06:00	1	-0.6	90.9%	0.0	0	100.6	8.8	2.3	
12	2012-12-31 06:00	2012-12-31 07:00	1	0.0	90.0%	0.0	0	100.5	8.9	2.5	
13	2012-12-31 07:00	2012-12-31 08:00	1	0.4	89.5%	0.0	0	100.5	9.1	2.6	
14	2012-12-31 08:00	2012-12-31 09:00	1	0.8	88.6%	0.0	37	100.4	9.5	2.8	
15	2012-12-31 09:00	2012-12-31 10:00	1	1.3	86.6%	0.0	125	100.3	10.4	3.2	
16	2012-12-31 10:00	2012-12-31 11:00	1	1.7	83.3%	0.0	206	100.3	11.1	3.6	
17	2012-12-31 11:00	2012-12-31 12:00	1	2.0	80.9%	0.0	228	100.1	11.1	3.7	
18	2012-12-31 12:00	2012-12-31 13:00	1	2.1	80.1%	0.0	188	100.0	11.0	3.7	

notes

Sunday, July 10, 2022 8:18 PM

covid

writing report

RETScreen - the data generated differs based on location selected, but when selecting around downtown Toronto, I was getting only about 2 different sets of results, so I'm guessing it's satellite data, but it's not like as accurate as just "click a spot on the map and get the exact weather data"

which neural networks have been done? Is it just ANNs? Has muzammil gotten any success with neural networks

- having trouble with the input shape values when I tried implementing LSTM

how should I proceed now

poster presentations - what time works best on sept 1

if time: try a different location like vancouver
try different location around toronto
try to remove solar radiation and see if I get good r^2

try gov data from prof
start finishing a write up
try diff stations again
check different lag times 1-3 empirically
find conferences?

notes jul 17

Sunday, July 17, 2022 7:17 PM

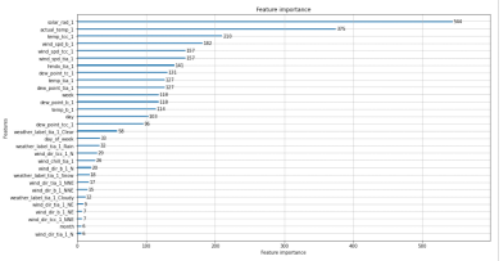
Note:
- can't use mape
- same with smape

summary:
- solar radiation helps a lot
- locations of weather stations plays a massive role in accuracy of model

results of models with and without solar radiation

solar

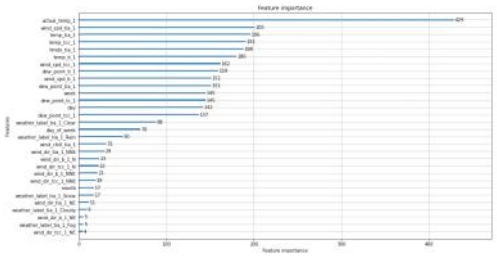
```
In [23]: train,test = data_load_pipeline(r"../data/processed_data/summer_day_df.csv",data_type="summer+day")

[LightGBM] [Warning] Auto-choosing col-wise multi-threading, the overhead of testing was 0.003868 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 3000
[LightGBM] [Info] Number of data points in the train set: 14195, number of used features: 32
[LightGBM] [Info] Start training from score 15.881825
```



summer+day
RMSE of LightGBM: 0.8839511405768282
MAE of LightGBM: 0.6549234307643715
R2 of LightGBM: 0.9909011548797781

no solar

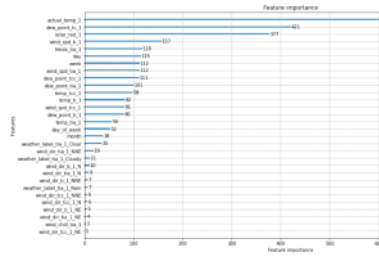
```
In [35]: train,test = data_load_pipeline(r"../data/processed_data/summer_day_df.csv",data_type="summer+day")

[LightGBM] [Warning] Auto-choosing col-wise multi-threading, the overhead of testing was 0.000643 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2751
[LightGBM] [Info] Number of data points in the train set: 14195, number of used features: 31
[LightGBM] [Info] Start training from score 15.881825
```



summer+day
RMSE of LightGBM: 0.9512915178499388
MAE of LightGBM: 0.7292805555032253
R2 of LightGBM: 0.9895270585389439

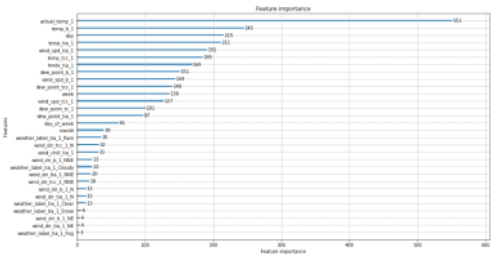
king city

```
In [7]: train,test = data_load_pipeline(r"../data/processed_data/summer_day_df_king.csv",data_type="
[LightGBM] [Warning] Auto-choosing col-wise multi-threading, the overhead of testing was 0.000796
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2978
[LightGBM] [Info] Number of data points in the train set: 14195, number of used features: 32
[LightGBM] [Info] Start training from score 9.163987
```



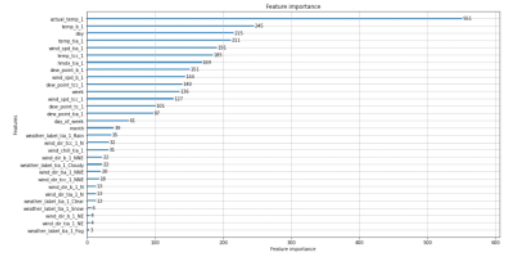
summer+day
RMSE of LightGBM: 1.0748284425722865
MAE of LightGBM: 0.7956364756200201
R2 of LightGBM: 0.9881492041881182

```
In [14]: train,test = data_load_pipeline(r"../data/processed_data/summer_night_df.csv",data_type="summer+night")

[LightGBM] [Warning] Auto-choosing col-wise multi-threading, the overhead of testing was 0.000865 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2746
[LightGBM] [Info] Number of data points in the train set: 7324, number of used features: 31
[LightGBM] [Info] Start training from score 12.733842
```



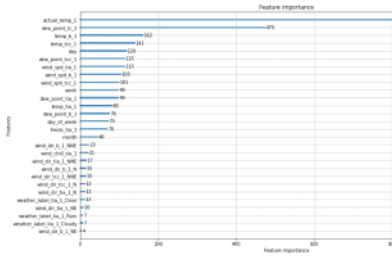
summer+night
RMSE of LightGBM: 0.6066747586778484
MAE of LightGBM: 0.4161781865965373
R2 of LightGBM: 0.9948258275922962

```
In [36]: train,test = data_load_pipeline(r"../data/processed_data/summer_night_df.csv",data_type="summer+night")

[LightGBM] [Warning] Auto-choosing col-wise multi-threading, the overhead of testing was 0.000712 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2746
[LightGBM] [Info] Number of data points in the train set: 7324, number of used features: 31
[LightGBM] [Info] Start training from score 12.733842
```



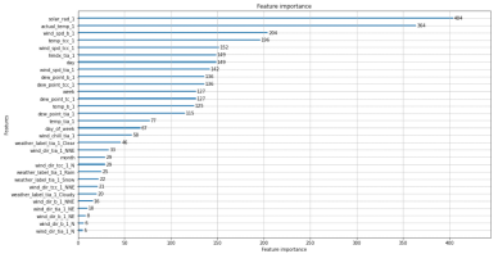
summer+night
RMSE of LightGBM: 0.6066747586778484
MAE of LightGBM: 0.4161781865965373
R2 of LightGBM: 0.9948258275922962

```
In [8]: train,test = data_load_pipeline(r"../data/processed_data/summer_night_df_king.csv",data_type="su
[LightGBM] [Warning] Auto-choosing col-wise multi-threading, the overhead of testing was 0.000079 sec
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2674
[LightGBM] [Info] Number of data points in the train set: 7324, number of used features: 31
[LightGBM] [Info] Start training from score 0.189104
```



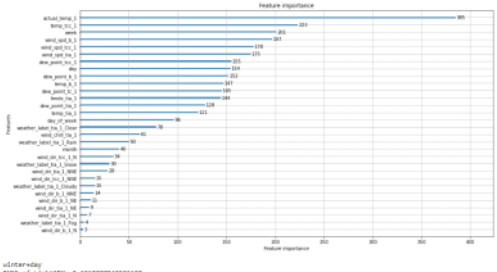
summer+night
RMSE of LightGBM: 0.7546765596260915
MAE of LightGBM: 0.5868721653660653
R2 of LightGBM: 0.9926350627728044432

```
In [15]: train,test = data_load_pipeline(r"../data/processed_data/winter_day_df.csv",data_type="winter+day")

[LightGBM] [Warning] Auto-choosing col-wise multi-threading, the overhead of testing was 0.000875 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 3005
[LightGBM] [Info] Number of data points in the train set: 7673, number of used features: 32
[LightGBM] [Info] Start training from score 0.782002
```



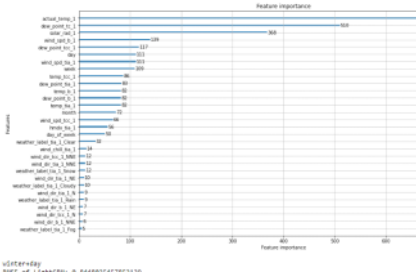
winter+day
RMSE of LightGBM: 0.6593575245327227
MAE of LightGBM: 0.4047363005366428
R2 of LightGBM: 0.9956551919100014

```
In [37]: train,test = data_load_pipeline(r"../data/processed_data/winter_day_df.csv",data_type="winter+day")

[LightGBM] [Warning] Auto-choosing col-wise multi-threading, the overhead of testing was 0.000848 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2751
[LightGBM] [Info] Number of data points in the train set: 7673, number of used features: 31
[LightGBM] [Info] Start training from score 0.782002
```



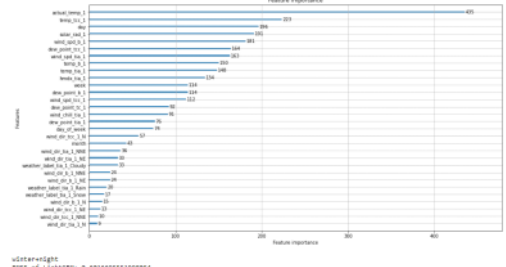
winter+day
RMSE of LightGBM: 0.6867777040683187
MAE of LightGBM: 0.5153378074675717
R2 of LightGBM: 0.9952860250738257

```
In [9]: train,test = data_load_pipeline(r"../data/processed_data/winter_day_df_king.csv",data_type="winter+day"
[LightGBM] [Warning] Auto-choosing col-wise multi-threading, the overhead of testing was 0.000920 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 3005
[LightGBM] [Info] Number of data points in the train set: 7673, number of used features: 32
[LightGBM] [Info] Start training from score 3.004917
```



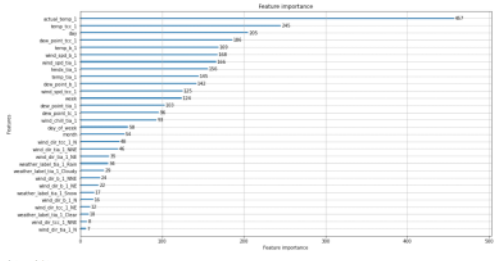
winter+day
RMSE of LightGBM: 0.8448021457852129
MAE of LightGBM: 0.6161562978373504
R2 of LightGBM: 0.9934760273496005

```
In [16]: train,test = data_load_pipeline(r"../data/processed_data/winter_night_df.csv",data_type="winter+night")

[LightGBM] [Warning] Auto-choosing col-wise multi-threading, the overhead of testing was 0.001025 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 3006
[LightGBM] [Info] Number of data points in the train set: 13896, number of used features: 32
[LightGBM] [Info] Start training from score 4.807266
```



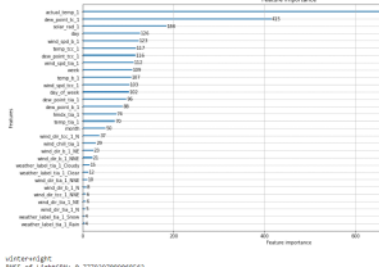
winter+night
RMSE of LightGBM: 0.6016055551008854
MAE of LightGBM: 0.3838631847907162
R2 of LightGBM: 0.9958400038203128

```
In [38]: train,test = data_load_pipeline(r"../data/processed_data/winter_night_df.csv",data_type="winter+night")

[LightGBM] [Warning] Auto-choosing col-wise multi-threading, the overhead of testing was 0.000884 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2751
[LightGBM] [Info] Number of data points in the train set: 13896, number of used features: 31
[LightGBM] [Info] Start training from score 4.997200
```



winter+night
RMSE of LightGBM: 0.6297958167376451
MAE of LightGBM: 0.4089749623037000
R2 of LightGBM: 0.9954531582083441

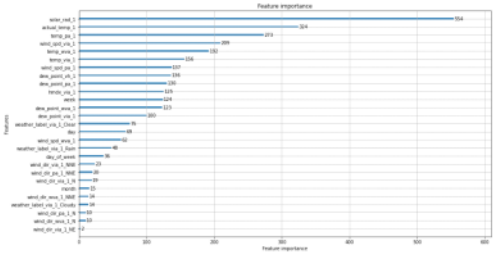
```
In [10]: train,test = data_load_pipeline(r"../data/processed_data/winter_night_df_king.csv",data_type
[LightGBM] [Warning] Auto-choosing col-wise multi-threading, the overhead of testing was 0.000874
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 3000
[LightGBM] [Info] Number of data points in the train set: 13896, number of used features: 32
[LightGBM] [Info] Start training from score 3.158000
```



winter+night
RMSE of LightGBM: 0.7779197000000562
MAE of LightGBM: 0.4959331002155865
R2 of LightGBM: 0.9958311588898526

van

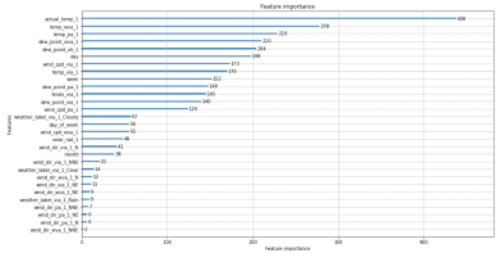
```
In [9]: 1 train,test = data_load_pipeline(r"../data/processed_data/summer_day_df_van.csv",data_type="summer+day")

[LightGBM] [Warning] Auto-choosing col-wise multi-threading, the overhead of testing was 0.001873 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2091
[LightGBM] [Info] Number of data points in the train set: 14233, number of used features: 31
[LightGBM] [Info] Start training from score 15.486070
```




```
summer+day
RMSE of LightGBM: 0.9014885854780044
MAE of LightGBM: 0.5940593485724824
R2 of LightGBM: 0.9816951847783669
```

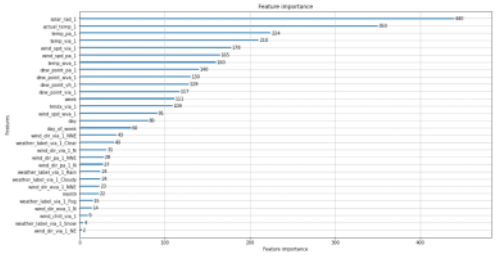
-

```
In [10]: 1 train,test = data_load_pipeline(r"../data/processed_data/summer_night_df_van.csv",data_type="summer+night")

[LightGBM] [Warning] Auto-choosing col-wise multi-threading, the overhead of testing was 0.002280 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2167
[LightGBM] [Info] Number of data points in the train set: 7148, number of used features: 31
[LightGBM] [Info] Start training from score 12.377957
```



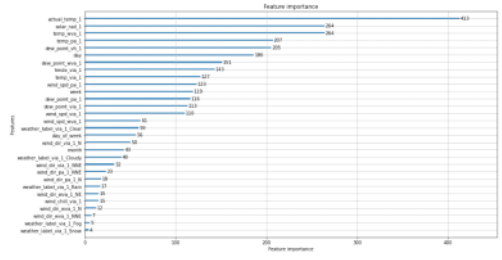

```
summer+night
RMSE of LightGBM: 0.6469117191366816
MAE of LightGBM: 0.4675280745026675
R2 of LightGBM: 0.9838411211703266
```

```
In [11]: 1 train,test = data_load_pipeline(r"../data/processed_data/winter_day_df_van.csv",data_type="winter+day")

[LightGBM] [Warning] Auto-choosing col-wise multi-threading, the overhead of testing was 0.000679 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2705
[LightGBM] [Info] Number of data points in the train set: 7229, number of used features: 32
[LightGBM] [Info] Start training from score 9.630332
```




```
winter+day
RMSE of LightGBM: 0.6544014351655626S
MAE of LightGBM: 0.4508819412611143
R2 of LightGBM: 0.9790691160005198
```

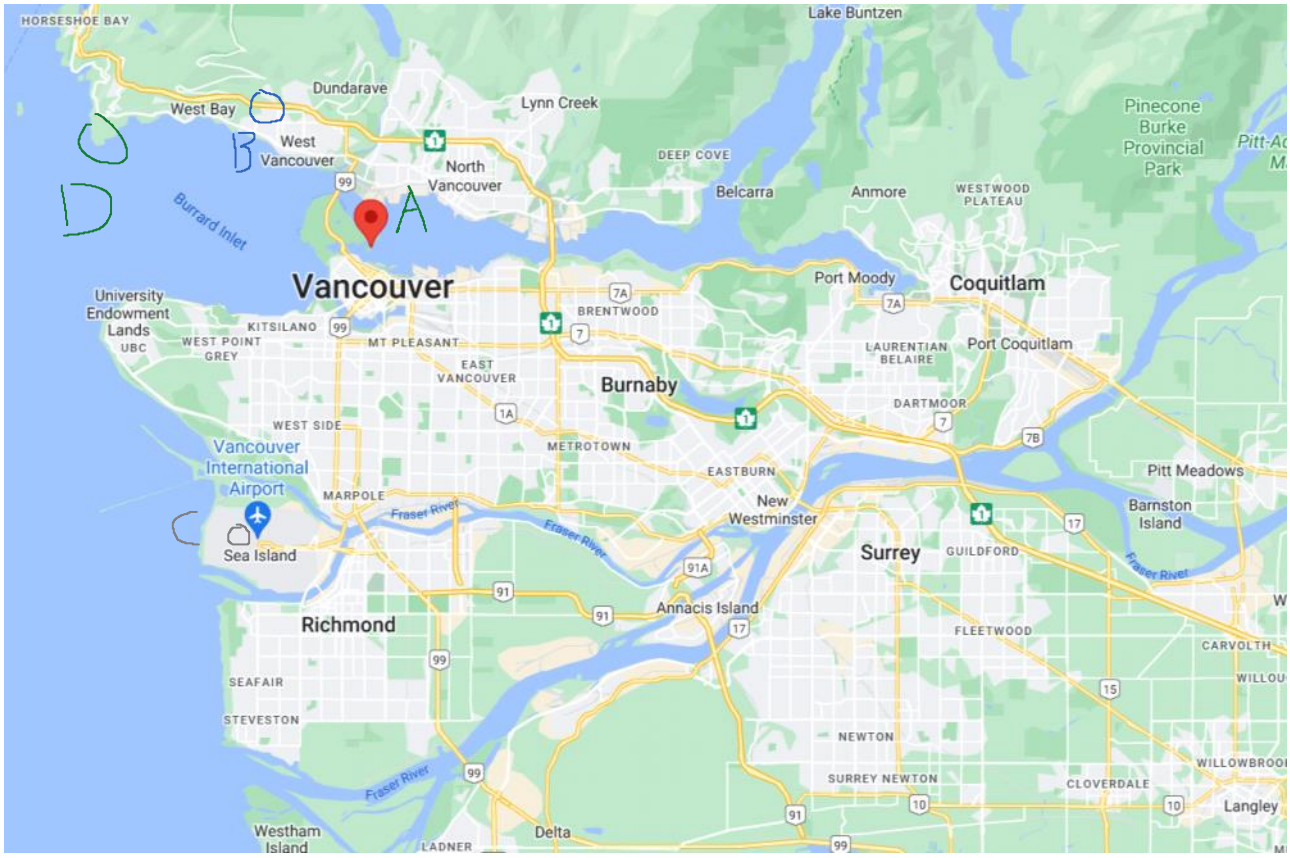
```
In [12]: 1 train,test = data_load_pipeline(r"../data/processed_data/winter_night_df_van.csv",data_type="winter+night")

[LightGBM] [Warning] Auto-choosing col-wise multi-threading, the overhead of testing was 0.000726 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2646
[LightGBM] [Info] Number of data points in the train set: 12065, number of used features: 32
[LightGBM] [Info] Start training from score 7.841716
```




```
winter+night
RMSE of LightGBM: 0.5735609336463019
MAE of LightGBM: 0.3925006280985025
R2 of LightGBM: 0.9881652658141463
```

notes jul 18

Monday, July 18, 2022 12:08 AM



A = vancouver harbour CS
B = West Vancouver Aut
C = vancouver intl a
D = point atkinson