6210 Assignment1 README

1. Firstly I find dataset through KAGGLE, this is a climate records about New York Daily Climate change.
2. I download this dataset and use python to cleaning data on jupyter notebook.
3. I check if there are some null value on dataset, and use the word to fill the blank.
4. I also notice that the value of Wind Speed is not correct in some columns, it' s not reasonable that the Wind Speed is 0, so I use the average of all rows' wind speed to fill those wrong columns. Then I check the value of attributes' min and max, and I notice some of the sea level pressure is lower than 0, and I also use average to fill them.
5. Then I check if this dataset exist some duplicate rows, and I also use curve chart to analyze the consistency and uniformity of some attributes.
6. After cleaning data, I use python to inset three tables to MySQL database.
7. 5 us-cases

   1>Join table1 and table3, then search temperature when events are snow and show 20 results:

```
SELECT wendi1.TemperatureC
FROM wendi1 INNER JOIN w3
ON wendi1.TimeID = w3.TimeID
WHERE w3.Events = "Snow"
ORDER BY wendi1.TemperatureC DESC
LIMIT 20 OFFSET 10
```

+ 选项

TemperatureC 1
1.1
1
1
1
0.6
0.6
0.6
0.6
0.6
0.6
0.6
0.6
0.6
0.6
0.6
0.6
0
0
0

2>Search dew point and humidity

```
SELECT wendi1.DewPointC,wendi1.Humidity
FROM wendi1 INNER JOIN w2
ON wendi1.TimeID = w2.TimeID
WHERE w2.VisibilityKm < 5
ORDER BY wendi1.Humidity ASC
LIMIT 10 OFFSET 5;
```

+ 选项

DewPointC	Humidity ▲ 1
−10.6	65
−4	65
−8.3	66
−8.9	66
−3.9	67
−8.3	69
−8.3	69
−8.3	69
−7.8	71
−7.8	71

3>

```
SELECT w2.TimeID,w2.SeaLevelPressurehPa,w3.Precipitationmm
FROM w2 INNER JOIN w3
ON w2.TimeID = w3.TimeID
WHERE w3.Precipitationmm > 0
ORDER BY w2.TimeID ASC
LIMIT 15
```

+ 选项

TimeID ▲ 1	SeaLevelPressurehPa	Precipitationmm
235	1012.40000	0.04
236	1011.70000	0.05
237	1011.70000	0.05
238	1010.70000	0.14
239	1010.40000	0.14
240	1009.80000	0.20
241	1010.40000	0.07
242	1008.40000	0.26
243	1009.00000	0.29
244	1007.80000	0.32
245	1007.70000	0.04
246	1006.70000	0.12
247	1006.30000	0.19
248	1005.40000	0.20
249	1002.20000	0.47

4>

```
SELECT w2.TimeID,w2.WindSpeedKmh,w2.GustSpeedKmh,w3.WindDirDegrees
FROM w2 INNER JOIN w3
ON w2.TimeID = w3.TimeID
WHERE w3.WindDirDegrees > 100
ORDER BY w2.TimeID DESC
LIMIT 20 OFFSET 20
```

+ 选项

TimeID ▾ 1	WindSpeedKmh	GustSpeedKmh	WindDirDegrees
4439	5.6	0.0	250
4437	5.6	0.0	280
4432	7.4	0.0	160
4406	5.6	0.0	120
4367	5.6	0.0	360
4354	9.3	33.3	300
4353	13.0	0.0	320
4352	18.5	29.6	300
4350	14.8	31.5	290
4349	20.4	33.3	290
4348	16.7	29.6	270
4346	11.1	0.0	280
4339	5.6	0.0	130
4326	16.7	0.0	290
4322	13.0	0.0	300
4320	5.6	0.0	230
4318	9.3	0.0	260
4315	14.8	0.0	260
4314	13.0	31.5	260
4304	9.3	0.0	170

5>

```
SELECT wendi1.TimeID,wendi1.DewPointC, w2.SeaLevelPressurehPa,w2.VisibilityKm
FROM wendi1 JOIN w2
on wendi1.TimeID = w2.TimeID
WHERE w2.SeaLevelPressurehPa=(SELECT MIN(w2.SeaLevelPressurehPa) FROM w2)
```

8. SQL Schema that makes sense and is in at least in first normal form

w2

Column	Type	Null	Default	Comments
TimeID	int(4)	Yes	*NULL*	
SeaLevelPressurehPa	decimal(9,5)	Yes	*NULL*	
VisibilityKm	decimal(3,1)	Yes	*NULL*	
WindDirection	varchar(8)	Yes	*NULL*	
WindSpeedKmh	decimal(4,1)	Yes	*NULL*	
GustSpeedKmh	decimal(3,1)	Yes	*NULL*	

w3

Column	Type	Null	Default	Comments
TimeID	int(4)	Yes	*NULL*	
Precipitationmm	decimal(3,2)	Yes	*NULL*	
Events	varchar(8)	Yes	*NULL*	
Conditions	varchar(19)	Yes	*NULL*	
WindDirDegrees	int(3)	Yes	*NULL*	

wendi1

Column	Type	Null	Default	Comments
TimeID	int(11)	Yes	*NULL*	
TemperatureC	float	Yes	*NULL*	
DewPointC	float	Yes	*NULL*	
Humidity	int(11)	Yes	*NULL*	