

Part 1 Python Data Cleaning

NY Weather Data Apr 14- Jul 14.xlsx

2014 Temp. (°C)			Dew Point (°C)			Humidity (%)			Sea Level Press. (hPa)			Visibility (km)			Wind (km/h)			Precip. (mm)	Events			
lpr	high	avg	low	high	avg	low	high	avg	low	high	avg	low	high	avg	low	high	avg	low	high	sum		
1	2	14	8	3	2	-2	-4	76	56	35	1021	1019	1016	16	16	16	26	16	40		0	
2	9	9	6	3	5	3	1	92	81	70	1022	1020	1019	16	16	16	16	6	32	0.76	Rain	
3	3	17	11	4	5	1	-6	92	57	22	1021	1019	1017	16	16	13	21	9	26	2.03	Rain	
4	4	7	6	5	6	3	2	92	84	76	1019	1015	1007	16	13	4	42	22	52	2.54	Rain	
5	5	14	9	5	6	0	-6	93	63	32	1019	1011	1007	16	14	3	53	27	63	T	Rain	
6	6	14	8	3	2	-4	-8	79	50	21	1024	1022	1019	16	16	16	40	17	89		0	
7	7	11	7	3	8	4	0	93	74	54	1024	1017	1007	16	11	3	42	18	56	8.89	Rain	
8	8	19	14	8	13	7	-1	93	65	37	1006	1000	996	16	7	0	48	22	60	6.35	Fog , Rain , Thunderstorm	
9	9	18	12	6	1	-3	-12	65	43	20	1017	1008	1004	16	16	16	48	24	60		0	
10	10	12	8	3	4	-2	-13	71	46	21	1024	1021	1018	16	16	16	47	26	56		0	
11	11	21	14	8	9	6	3	86	61	35	1018	1016	1014	16	16	16	32	14	39	T	Rain	
12	12	20	15	9	10	6	1	83	56	29	1024	1022	1016	16	16	16	29	13	39		0	
13	13	18	13	8	11	9	6	89	75	60	1023	1020	1016	16	16	13	50	24	61		0	
14	14	19	15	11	12	10	9	93	77	60	1018	1015	1012	16	15	1	58	36	76		0	
15	15	13	7	1	12	9	0	96	88	79	1012	1007	1001	16	5	0	53	37	69	16.51	Fog , Rain , Snow	
16	16	9	5	0	0	-7	-11	64	44	23	1035	1025	1013	16	15	8	50	26	60	2.03	Rain , Snow	
17	17	8	5	2	-1	-4	-8	70	55	40	1041	1038	1036	16	16	16	35	20	52		0	
18	18	8	5	2	2	-2	-4	79	64	49	1037	1032	1026	16	16	16	48	11	232		0	
19	19	20	12	4	3	-1	-8	85	51	16	1028	1025	1024	16	16	16	32	16	113		0	
20	20	12	9	6	1	-3	-6	60	44	28	1033	1031	1028	16	16	16	35	19	42		0	
21	21	14	9	3	6	1	-4	86	58	30	1027	1020	1013	16	16	16	21	10	24		0	
22	22	17	12	7	12	8	6	93	77	60	1012	1005	1000	16	16	14	39	14	47	T	Rain	
23	23	17	12	7	9	2	-3	83	59	35	1010	1003	999	16	16	16	60	36	76		0	
24	24	18	12	6	-2	-10	-16	53	32	10	1017	1012	1010	16	16	16	50	34	66		0	
25	25	15	11	6	5	-3	-9	77	50	22	1018	1015	1009	16	16	16	29	15	37	T	Rain	
26	26	19	14	8	8	7	4	93	65	37	1012	1008	1006	16	14	2	35	17	42	18.03	Rain	
27	27	16	12	8	7	1	-1	86	60	33	1020	1015	1012	16	16	16	37	22	50		0	
28	28	19	13	7	6	2	-4	74	48	22	1021	1019	1016	16	16	16	35	16	40		0	
29	29	11	9	7	5	3	2	77	69	61	1030	1027	1022	16	16	16	50	32	63	1.27	Rain	
30	30	13	9	6	12	6	3	100	88	76	1029	1024	1015	16	6	2	50	31	63	118.87	Rain	
2014 Temp. (°C)			Dew Point (°C)			Humidity (%)			Sea Level Press. (hPa)			Visibility (km)			Wind (km/h)			Precip. (mm)	Events			
day	high	avg	low	high	avg	low	high	avg	low	high	avg	low	high	avg	low	high	avg	low	high	sum		
1	2	22	17	12	16	12	11	100	89	78	1015	1009	1006	16	7	0	37	21	48		6.1	Fog , Rain
2	2	21	16	11	11	6	0	72	51	30	1011	1008	1006	16	16	11	39	18	45	0.25	Rain	
3	3	19	14	9	10	6	4	89	65	40	1012	1010	1008	16	16	16	37	16	45	T	Rain	
4	4	19	16	12	9	3	-6	83	53	22	1009	1007	1004	16	16	16	52	30	71	T		
5	5	21	16	10	2	-2	-8	54	35	15	1012	1011	1009	16	16	16	53	27	69		0	
6	6	21	16	10	6	-1	-5	51	37	22	1019	1014	1012	16	16	16	35	17	47		0	
7	7	18	14	10	8	1	-4	77	51	25	1024	1023	1019	16	16	16	32	15	37		0	

Cleaned data

Date	Temp. (°C)-high	Temp. (°C)-avg	Temp. (°C)-low	Dew Point (°C)-high	Dew Point (°C)-avg	Dew Point (°C)-low	Humidity (%) -high	Humidity (%) -avg	Humidity (%) -low	Sea Level Press. (hPa)-high	Sea Level Press. (hPa)-avg
2014/4/1	14	8	3	2	-2	-4	76	56	35	1021	1019
2014/4/2	9	6	3	5	3	1	92	81	70	1022	1020
2014/4/3	17	11	4	5	1	-6	92	57	22	1021	1019
2014/4/4	7	6	5	6	3	2	92	84	76	1019	1015
2014/4/5	14	9	5	6	0	-6	93	63	32	1019	1011
2014/4/6	14	8	3	2	-4	-8	79	50	21	1024	1022
2014/4/7	11	7	3	8	4	0	93	74	54	1024	1017
2014/4/8	19	14	8	13	7	-1	93	65	37	1006	1000
2014/4/9	18	12	6	1	-3	-12	65	43	20	1017	1008
2014/4/10	12	8	3	4	-2	-13	71	46	21	1024	1021
2014/4/11	21	14	8	9	6	3	86	61	35	1018	1016
2014/4/12	20	15	9	10	6	1	83	56	29	1024	1022
2014/4/13	18	13	8	11	9	6	89	75	60	1023	1020
2014/4/14	19	15	11	12	10	9	93	77	60	1018	1015
2014/4/15	13	7	1	12	9	0	96	88	79	1012	1007
2014/4/16	9	5	0	0	-7	-11	64	44	23	1035	1025
2014/4/17	8	5	2	-1	-4	-8	70	55	40	1041	1038
2014/4/18	8	5	2	2	-2	-4	79	64	49	1037	1032
2014/4/19	20	12	4	3	-1	-8	85	51	16	1028	1025
2014/4/20	12	9	6	1	-3	-6	60	44	28	1033	1031
2014/4/21	14	9	3	6	1	-4	86	58	30	1027	1020
2014/4/22	17	12	7	12	8	6	93	77	60	1012	1005
2014/4/23	17	12	7	9	2	-3	83	59	35	1010	1003
2014/4/24	18	12	6	-2	-10	-16	53	32	10	1017	1012
2014/4/25	15	11	6	5	-3	-9	77	50	22	1018	1015
2014/4/26	19	14	8	8	7	4	93	65	37	1012	1008
2014/4/27	16	12	8	7	1	-1	86	60	33	1020	1015
2014/4/28	19	13	7	6	2	-4	74	48	22	1021	1019
2014/4/29	11	9	7	5	3	2	77	69	61	1030	1027
2014/4/30	13	9	6	12	6	3	100	88	76	1029	1024
2014/5/1	22	17	12	16	12	11	100	89	78	1015	1009
2014/5/2	21	16	11	11	6	0	72	51	30	1011	1008

Step 1: Raw Data (Original Excel File)

2014	Temp	Unnamed:2	Unnamed:3	Rainfall	Unnamed:5	Unnamed:6
Apr	Max	Avg	Min	Total	Avg	Max
1	20	15	10	0.0	0.0	0.0
2	22	16	11	0.2	0.1	0.2
May	Max	Avg	Min	Total	Avg	Max
1	25	19	13	0.0	0.0	0.0
2	24	18	12	0.1	0.1	0.1

Step 2: After Identifying Month Headers and Splitting into Separate Monthly DataFrames

- April DataFrame:

2014	Temp-Max	Temp-Avg	Temp-Min	Rainfall-Total	Rainfall-Avg	Rainfall-Max
Apr	Max	Avg	Min	Total	Avg	Max
1	20	15	10	0.0	0.0	0.0
2	22	16	11	0.2	0.1	0.2

- May DataFrame:

2014	Temp-Max	Temp-Avg	Temp-Min	Rainfall-Total	Rainfall-Avg	Rainfall-Max
May	Max	Avg	Min	Total	Avg	Max
1	25	19	13	0.0	0.0	0.0
2	24	18	12	0.1	0.1	0.1

Step 3: After Formatting Dates and Standardizing Columns

- April DataFrame (Cleaned):

Date	Temp-Max	Temp-Avg	Temp-Min	Rainfall-Total	Rainfall-Avg	Rainfall-Max
2014/4/1	20	15	10	0.0	0.0	0.0
2014/4/2	22	16	11	0.2	0.1	0.2

- May DataFrame (Cleaned):

Date	Temp-Max	Temp-Avg	Temp-Min	Rainfall-Total	Rainfall-Avg	Rainfall-Max
2014/5/1	25	19	13	0.0	0.0	0.0
2014/5/2	24	18	12	0.1	0.1	0.1

Step 4: Final Combined Dataset

Date	Temp-Max	Temp-Avg	Temp-Min	Rainfall-Total	Rainfall-Avg	Rainfall-Max
2014/4/1	20	15	10	0.0	0.0	0.0
2014/4/2	22	16	11	0.2	0.1	0.2
2014/5/1	25	19	13	0.0	0.0	0.0
2014/5/2	24	18	12	0.1	0.1	0.1

Part 2 Build Worktables using SQL

Date	Temp. (°C)-high	Temp. (°C)-avg	Temp. (°C)-low	Dew Point (°C)-high	Dew Point (°C)-avg	Dew Point (°C)-low	Humidity (%) -high	Humidity (%) -avg	Humidity (%) -low	Sea Level Press. (hPa)-high	Sea Level Press. (hPa)-avg
2014/4/1	14	8	3	2	-2	-4	76	56	35	1021	1019
2014/4/2	9	6	3	5	3	1	92	81	70	1022	1020
2014/4/3	17	11	4	5	1	-6	92	57	22	1021	1019
2014/4/4	7	6	5	6	3	2	92	84	76	1019	1015
2014/4/5	14	9	5	6	0	-6	93	63	32	1019	1011
2014/4/6	14	8	3	2	-4	-8	79	50	21	1024	1022
2014/4/7	11	7	3	8	4	0	93	74	54	1024	1017
2014/4/8	19	14	8	13	7	-1	93	65	37	1006	1000
2014/4/9	18	12	6	1	-3	-12	65	43	20	1017	1008
2014/4/10	12	8	3	4	-2	-13	71	46	21	1024	1021
2014/4/11	21	14	8	9	6	3	86	61	35	1018	1016
2014/4/12	20	15	9	10	6	1	83	56	29	1024	1022
2014/4/13	18	13	8	11	9	6	89	75	60	1023	1020
2014/4/14	19	15	11	12	10	9	93	77	60	1018	1015
2014/4/15	13	7	1	12	9	0	96	88	79	1012	1007
2014/4/16	9	5	0	0	-7	-11	64	44	23	1035	1025
2014/4/17	8	5	2	-1	-4	-8	70	55	40	1041	1038
2014/4/18	8	5	2	2	-2	-4	79	64	49	1037	1032
2014/4/19	20	12	4	3	-1	-8	85	51	16	1028	1025
2014/4/20	12	9	6	1	-3	-6	60	44	28	1033	1031
2014/4/21	14	9	3	6	1	-4	86	58	30	1027	1020
2014/4/22	17	12	7	12	8	6	93	77	60	1012	1005
2014/4/23	17	12	7	9	2	-3	83	59	35	1010	1003
2014/4/24	18	12	6	-2	-10	-16	53	32	10	1017	1012
2014/4/25	15	11	6	5	-3	-9	77	50	22	1018	1015
2014/4/26	19	14	8	8	7	4	93	65	37	1012	1008
2014/4/27	16	12	8	7	1	-1	86	60	33	1020	1015
2014/4/28	19	13	7	6	2	-4	74	48	22	1021	1019
2014/4/29	11	9	7	5	3	2	77	69	61	1030	1027
2014/4/30	13	9	6	12	6	3	100	88	76	1029	1024
2014/5/1	22	17	12	16	12	11	100	89	78	1015	1009
2014/5/2	21	16	11	11	6	0	72	51	30	1011	1008

uber-raw-data-apr14

Date/Time	Lat	Lon	Base
4/1/2014 0:11:00	40.769	-73.9549	B02512
4/1/2014 0:17:00	40.7267	-74.0345	B02512
4/1/2014 0:21:00	40.7316	-73.9873	B02512
4/1/2014 0:28:00	40.7588	-73.9776	B02512
4/1/2014 0:33:00	40.7594	-73.9722	B02512
4/1/2014 0:33:00	40.7383	-74.0403	B02512
4/1/2014 0:39:00	40.7223	-73.9887	B02512
4/1/2014 0:45:00	40.762	-73.979	B02512
4/1/2014 0:55:00	40.7524	-73.996	B02512
4/1/2014 1:01:00	40.7575	-73.9846	B02512

uber-raw-data-jul14

Date/Time	Lat	Lon	Base
7/1/2014 0:03:00	40.7586	-73.9706	B02512
7/1/2014 0:05:00	40.7605	-73.9994	B02512
7/1/2014 0:06:00	40.732	-73.9999	B02512
7/1/2014 0:09:00	40.7635	-73.9793	B02512
7/1/2014 0:20:00	40.7204	-74.0047	B02512
7/1/2014 0:35:00	40.7487	-73.9869	B02512
7/1/2014 0:57:00	40.7444	-73.9961	B02512
7/1/2014 0:58:00	40.7132	-73.9492	B02512
7/1/2014 1:04:00	40.759	-73.973	B02512
7/1/2014 1:08:00	40.7601	-73.9823	B02512

uber-raw-data-aug14

Date/Time	Lat	Lon	Base
8/1/2014 0:03:00	40.7366	-73.9906	B02512
8/1/2014 0:09:00	40.726	-73.9918	B02512
8/1/2014 0:12:00	40.7209	-74.0507	B02512
8/1/2014 0:12:00	40.7387	-73.9856	B02512
8/1/2014 0:12:00	40.7323	-74.0077	B02512
8/1/2014 0:13:00	40.7349	-74.0033	B02512
8/1/2014 0:15:00	40.7279	-73.9542	B02512
8/1/2014 0:17:00	40.721	-73.9937	B02512
8/1/2014 0:19:00	40.7195	-74.006	B02512
8/1/2014 0:20:00	40.7448	-73.9799	B02512

Uber_lat_lon

base	ride_count	month	avg_lat	avg_lon
B02512	35536	April	40.745492880648200	-73.9813446221401
B02598	183263	April	40.73981247711750	-73.97634237970960
B02617	108001	April	40.73838399117320	-73.97652647016470
B02682	227808	April	40.74010496900430	-73.9765244999165
B02764	9908	April	40.73926620883810	-73.9792396374465
B02512	35021	July	40.743752198388	-73.97856740994290
B02598	245597	July	40.738596625918600	-73.97178016657970
B02617	310160	July	40.739007625562600	-73.97241355417810
B02682	196754	July	40.73905466871270	-73.97184899523240
B02764	8589	July	40.74270142219800	-73.97271130706510
B02512	31472	August	40.740744337553200	-73.97705486745600
B02598	220129	August	40.73732285722490	-73.97009346896920
B02617	355803	August	40.737795218416000	-73.97042876710260
B02682	173280	August	40.737899014824300	-73.96972035146490
B02764	48591	August	40.7374076156332	-73.96560053000970

hourly_rides

ride_date	hour_of_day	ride_count	month
2014-04-01	0	138	April
2014-04-01	1	66	April
2014-04-01	2	53	April
2014-04-01	3	93	April
2014-04-01	4	166	April
2014-04-01	5	333	April
2014-04-01	6	722	April
2014-04-01	7	1001	April
2014-04-01	8	780	April
2014-04-01	9	532	April
2014-04-01	10	477	April
2014-04-01	11	547	April
2014-04-01	12	469	April
2014-04-01	13	591	April
2014-04-01	14	736	April
2014-04-01	15	967	April
2014-04-01	16	1152	April
2014-04-01	17	1262	April
2014-04-01	18	1122	April
2014-04-01	19	1018	April
2014-04-01	20	597	April
2014-04-01	21	878	April
2014-04-01	22	529	April
2014-04-01	23	317	April
2014-04-02	0	161	April
2014-04-02	1	82	April

Summarizes hourly ride counts for each month

