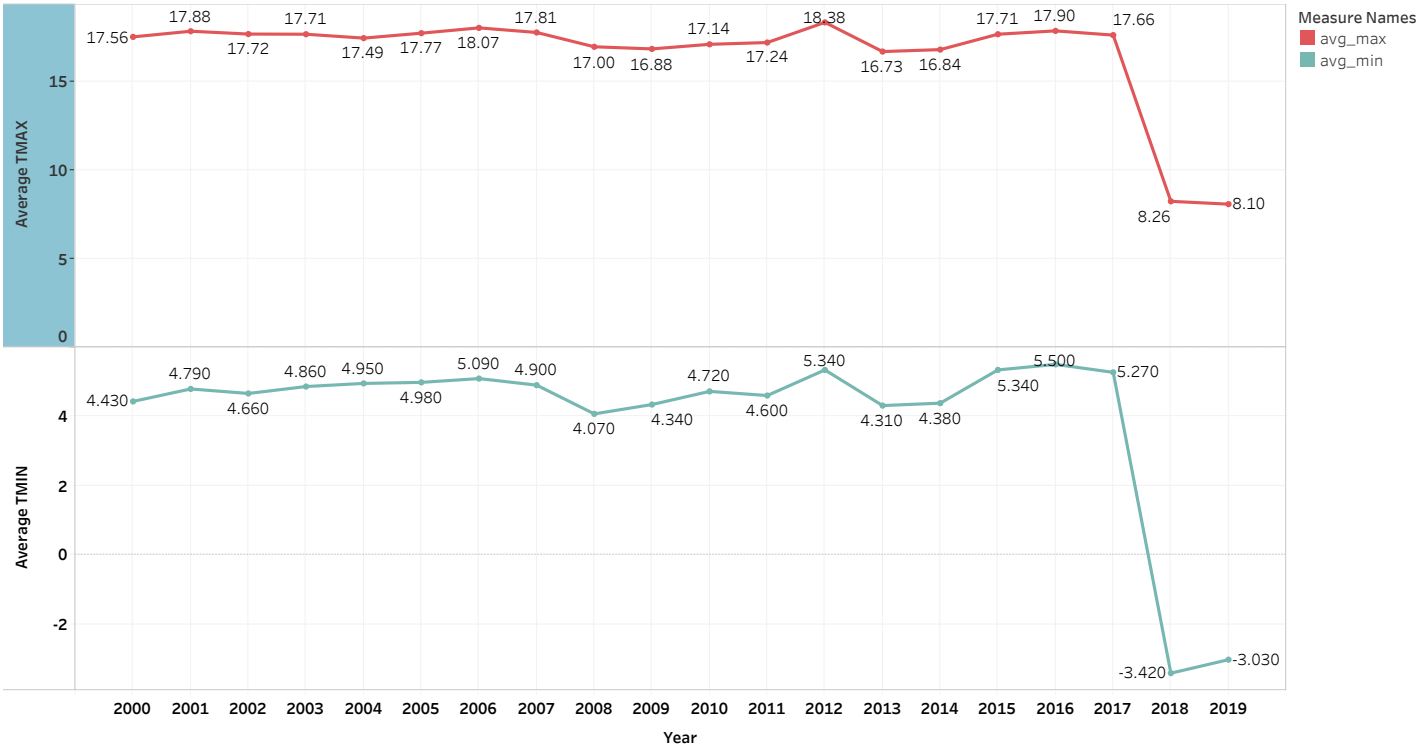
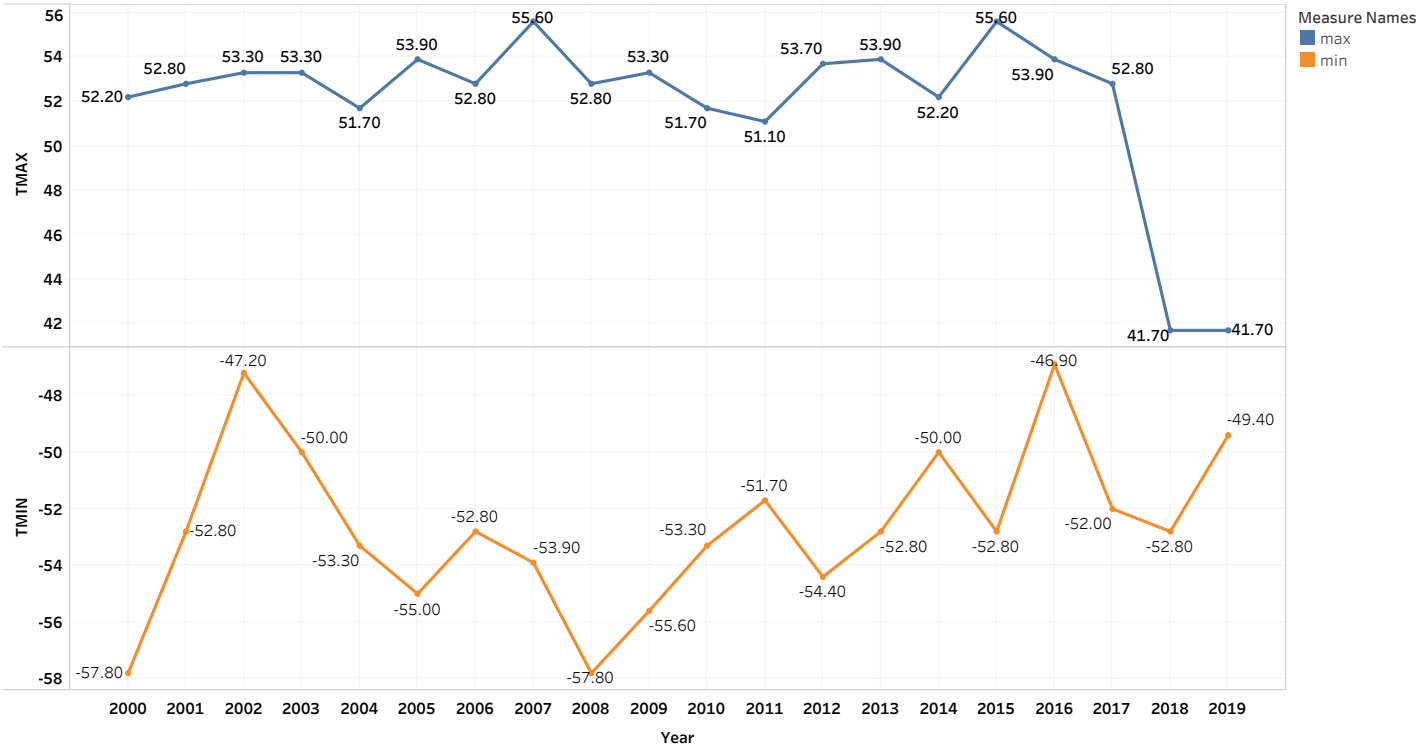
1.Average Tmax and Tmin over years



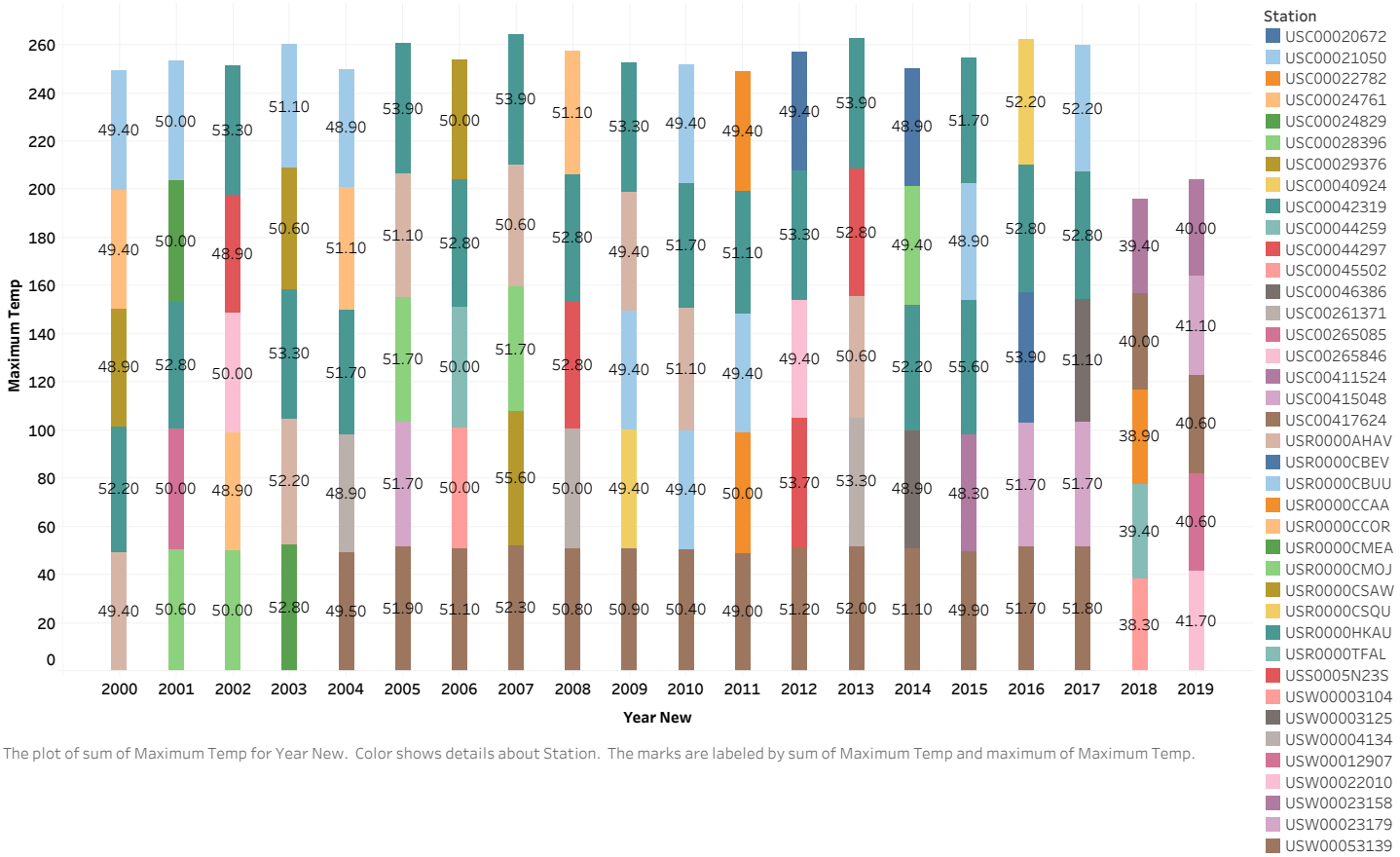
The trends of avg_max and avg_min for Year. Color shows details about avg_max and avg_min. The marks are labeled by sum of max and sum of min. For pane Sum of avg_max: The marks are labeled by sum of max, sum of min and avg_max. For pane Sum of avg_min: The marks are labeled by sum of max, sum of min and avg_min.

2.Maximum Tmax and Minimum Tmin over years



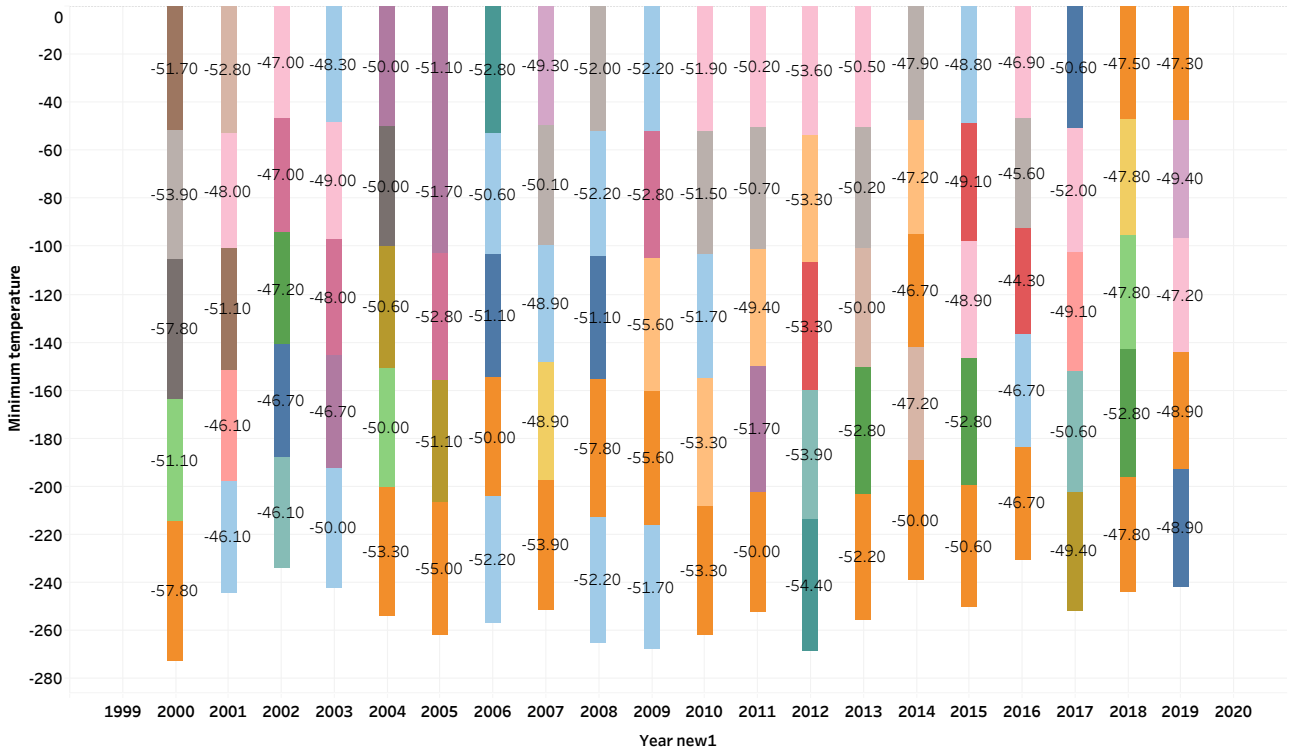
The trends of max and min for Year. Color shows details about max and min. The marks are labeled by max. For pane Sum of min: The marks are labeled by max and min.

3a) 5 Hottest stations for every year



The plot of sum of Maximum Temp for Year New. Color shows details about Station. The marks are labeled by sum of Maximum Temp and maximum of Maximum Temp.

3b) 5 Coldest stations for every year



The plot of sum of Minimum temperature for Year new1. Color shows details about Station11. The marks are labeled by sum of Minimum temperature.

1. Average TMIN, TMAX for each year excluding abnormalities or missing data

2. Maximum TMAX, Minimum TMIN for each year excluding abnormalities or missing data

Year	Average_Minimum Temp	Average_MaximumTemp	Minimum	Maximum
2000	4.43	17.56	-57.80	52.20
2001	4.79	17.88	-52.80	52.80
2002	4.66	17.72	-47.20	53.30
2003	4.86	17.71	-50.00	53.30
2004	4.95	17.49	-53.30	51.70
2005	4.98	17.77	-55.00	53.90
2006	5.09	18.07	-52.80	52.80
2007	4.90	17.81	-53.90	55.60
2008	4.07	17.00	-57.80	52.80
2009	4.34	16.88	-55.60	53.30
2010	4.72	17.14	-53.30	51.70
2011	4.60	17.24	-51.70	51.10
2012	5.34	18.38	-54.40	53.70
2013	4.31	16.73	-52.80	53.90
2014	4.38	16.84	-50.00	52.20
2015	5.34	17.71	-52.80	55.60
2016	5.50	17.90	-46.90	53.90
2017	5.27	17.66	-52.00	52.80
2018	-3.42	8.26	-52.80	41.70
2019	-3.03	8.10	-49.40	41.70

3. 5 hottest , 5 coldest weather stations for each year excluding abnormalities or missing data

Year	Station	Maximum_temperature
2000	USC00042319	52.20
2000	USC00024761	49.40
2000	USR0000AHAV	49.40
2000	USC00021050	49.40
2000	USC00029376	48.90
2001	USC00042319	52.80
2001	USR0000CMOJ	50.60
2001	USC00021050	50.00
2001	USC00024829	50.00
2001	USC00265085	50.00
2002	USC00042319	53.30
2002	USR0000CMOJ	50.00
2002	USC00265846	50.00
2002	USR0000CCOR	48.90
2002	USC00044297	48.90
2003	USC00042319	53.30
2003	USR0000CMEA	52.80
2003	USR0000AHAV	52.20
2003	USC00021050	51.10
2003	USC00029376	50.60
2004	USC00042319	51.70
2004	USC00024761	51.10
2004	USW00053139	49.50
2004	USC00021050	48.90
2004	USC00261371	48.90
2005	USC00042319	53.90

Year	Station	Minimum_temperature
2000	USC00505644	-57.80
2000	USC00501684	-57.80
2000	USC00508140	-53.90
2000	USW00026440	-51.70
2000	USC00502350	-51.10
2001	USW00026508	-52.80
2001	USR0000ABCA	-51.10
2001	USS0051R01S	-48.00
2001	USC00501492	-46.10
2001	USC00504567	-46.10
2002	USR0000AKAI	-47.20
2002	USS0051R01S	-47.00
2002	USS0050S01S	-47.00
2002	USR0000ABEV	-46.70
2002	USC00503212	-46.10
2003	USC00501492	-50.00
2003	USS0051R01S	-49.00
2003	USW00026533	-48.30
2003	USS0050S01S	-48.00
2003	USC00509869	-46.70
2004	USC00501684	-53.30
2004	USC00502568	-50.60
2004	USC00502350	-50.00
2004	USS0045O10S	-50.00
2004	USW00026412	-50.00
2005	USC00501684	-55.00

2005	USW00053139	51.90
2005	USR0000CMOJ	51.70
2005	USW00023179	51.70
2005	USR0000AHAV	51.10
2006	USC00042319	52.80
2006	USW00053139	51.10
2006	USC00029376	50.00
2006	USC00044259	50.00
2006	USC00045502	50.00
2007	USR0000CSAW	55.60
2007	USC00042319	53.90
2007	USW00053139	52.30
2007	USR0000CMOJ	51.70
2007	USR0000AHAV	50.60
2008	USC00042319	52.80
2008	USC00044297	52.80
2008	USC00024761	51.10
2008	USW00053139	50.80
2008	USC00261371	50.00
2009	USC00042319	53.30
2009	USW00053139	50.90
2009	USR0000CBUU	49.40
2009	USR0000CSQU	49.40
2009	USR0000AHAV	49.40
2010	USC00042319	51.70
2010	USR0000AHAV	51.10
2010	USW00053139	50.40
2010	USC00021050	49.40
2010	USR0000CBUU	49.40
2011	USC00042319	51.10
2011	USR0000CCAA	50.00
2011	USC00022782	49.40
2011	USR0000CBUU	49.40
2011	USW00053139	49.00
2012	USS0005N23S	53.70
2012	USC00042319	53.30
2012	USW00053139	51.20
2012	USC00020672	49.40
2012	USC00265846	49.40
2013	USC00042319	53.90
2013	USW00004134	53.30
2013	USC00044297	52.80
2013	USW00053139	52.00
2013	USR0000AHAV	50.60
2014	USC00042319	52.20
2014	USW00053139	51.10
2014	USC00028396	49.40
2014	USC00020672	48.90
2014	USW00003125	48.90
2015	USR0000HKAU	55.60
2015	USC00042319	51.70
2015	USW00053139	49.90
2015	USR0000CBUU	48.90
2015	USW00023158	48.30
2016	USR0000CBEV	53.90

2005	USC00509313	-52.80
2005	USC00509869	-51.70
2005	USC00502568	-51.10
2005	USW00026412	-51.10
2006	USR0000ASEL	-52.80
2006	USC00501492	-52.20
2006	USR0000ABEV	-51.10
2006	USR0000ACHL	-50.60
2006	USC00501684	-50.00
2007	USC00501684	-53.90
2007	USS0045R01S	-50.10
2007	USW00026422	-49.30
2007	USC00502607	-48.90
2007	USR0000ACHL	-48.90
2008	USC00501684	-57.80
2008	USC00501492	-52.20
2008	USR0000ACHL	-52.20
2008	USS0045R01S	-52.00
2008	USR0000ABEV	-51.10
2009	USC00502101	-55.60
2009	USC00501684	-55.60
2009	USC00509313	-52.80
2009	USR0000ACHL	-52.20
2009	USC00501492	-51.70
2010	USC00501684	-53.30
2010	USC00502101	-53.30
2010	USS0051R01S	-51.90
2010	USR0000ACHL	-51.70
2010	USS0045R01S	-51.50
2011	USC00509869	-51.70
2011	USS0045R01S	-50.70
2011	USS0051R01S	-50.20
2011	USC00501684	-50.00
2011	USR0000AFYK	-49.40
2012	USC00503165	-54.40
2012	USC00503212	-53.90
2012	USS0051R01S	-53.60
2012	USC00504210	-53.30
2012	USR0000AFYK	-53.30
2013	USC00502339	-52.80
2013	USC00501684	-52.20
2013	USS0051R01S	-50.50
2013	USS0045R01S	-50.20
2013	USR0000ABCK	-50.00
2014	USC00501684	-50.00
2014	USS0045R01S	-47.90
2014	USR0000ABCK	-47.20
2014	USR0000AFYK	-47.20
2014	USR0000ACOA	-46.70
2015	USC00502339	-52.80
2015	USC00501684	-50.60
2015	USS0041P07S	-49.10
2015	USC00509314	-48.90
2015	USW00026533	-48.80
2016	USS0051R01S	-46.90

2016	USC00042319	52.80
2016	USC00040924	52.20
2016	USW00023179	51.70
2016	USW00053139	51.70
2017	USC00042319	52.80
2017	USC00021050	52.20
2017	USW00053139	51.80
2017	USW00023179	51.70
2017	USC00046386	51.10
2018	USC00417624	40.00
2018	USC00411524	39.40
2018	USR0000TFAL	39.40
2018	USR0000CCAA	38.90
2018	USW00003104	38.30
2019	USW00022010	41.70
2019	USC00415048	41.10
2019	USC00417624	40.60
2019	USW00012907	40.60
2019	USC00411524	40.00

2016	USR0000ACHL	-46.70
2016	USC00501684	-46.70
2016	USS0045R01S	-45.60
2016	USS0041P07S	-44.30
2017	USS0051R01S	-52.00
2017	USR0000ASLC	-50.60
2017	USW00026529	-50.60
2017	USR0000ALIV	-49.40
2017	USS0045O04S	-49.10
2018	USC00502339	-52.80
2018	USC00501684	-47.80
2018	USR0000AKAV	-47.80
2018	USR0000ANOR	-47.80
2018	USW00096406	-47.50
2019	USC00509891	-49.40
2019	USC00211840	-48.90
2019	USC00501684	-48.90
2019	USW00096406	-47.30
2019	USC00509314	-47.20

4.Hottest and coldest day and corresponding weather stations in the entire dataset

Coldest	20080207	USC00501684	"-57.8C"
Hottest	20150213	USR0000HKAU	"55.6C"

Method Used : Hive SQL

Why Hive:

Hive gives an SQL-like interface to query data stored in various databases and file systems that integrate with **Hadoop. I** chose this because it has a very less learning curve and easy to implement.

Query 1 : SELECT substring(date_id,1,4)as year, measurement, AVG(value/10) as Average FROM Harini WHERE flag2 = '' GROUP BY measurement,substring(date_id,1,4) ORDER BY substring(date_id,1,4),measurement ;

Query 2a: SELECT substring(date_id,1,4) as year, measurement, MAX(value/10) as MAX_TMAX FROM Harini WHERE flag2 = '' and measurement = 'TMAX' GROUP BY measurement,substring(date_id,1,4) ORDER BY substring(date_id,1,4),measurement ;

Query 2b :SELECT substring(date_id,1,4) as year, measurement, MAX(value/10) as MAX_TMAX FROM Harini WHERE flag2 = '' and measurement = 'TMAX' GROUP BY measurement,substring(date_id,1,4) ORDER BY substring(date_id,1,4),measurement ;

Query 3a: SELECT station as MIN_station, VALUE, substr(date_id, 0, 4) AS YEAR, OVER (PARTITION BY SUBSTR(WDATE,0,4) ORDER BY VALUE) AS RANK FROM Harini WHERE FLAG2 = '' and ELEMENT = 'TMIN') AS MIN_VAL WHERE MIN_VAL.RANK <= 5

Query 3b: SELECT station as MAX_station, VALUE, substr(date_id, 0, 4) AS YEAR, OVER (PARTITION BY SUBSTR(WDATE,0,4) ORDER BY VALUE) AS RANK FROM Harini WHERE FLAG2 = ''' and ELEMENT = 'TMAX') AS MAX_VAL WHERE MAX_VAL.RANK <= 5

Query 4a: select * from Harini T where (T.value in (select max(value) from Harini where FLAG2="")) and FLAG2="";

Query 4b: select * from Harini T where (T.value in (select min(value) from Harini where FLAG2="")) and FLAG2="";