Report:

My data understanding:

1. **“Qualitaet_Niveau”**:Describes the quality control procedure applied for a data report for reporting time.
   1.1. Refers to a complete set of parameters at a specific date.
   1.2. Procedure can reach the best quality check level when Qualitaet_Niveau value is greater than or equal to 5(**Assumption**).
2. **“Wert”:**Is the temperature.
3. **Approach and Observations:**
   3.1. Filtered out the rows/instances whose values for column “**Qualitaet_Niveau**” measured is less than 5.
   3.2. Above step is done only for the years **2015, 2016, 2017, 2018** where there are sufficient observations to interpolate the series for a 15 minutes interval.
   3.3. Whereas for years **2019** and **2020** we can't interpolate as there are very few data points available for column Qualitaet_Niveau whose value is greater than equal to 5. So, for the year **2019** the number of instances present are 23 for column “**Qualitaet_Niveau”** for which values are 5,6,7.
   3.4. For **2020** count of instances are nil for columns **Qualitaet_Niveau** whose values are 5,6,7 and with **Qualitaet_Niveau :0** count of instances are 1345.
   3.5. Filtered values are then grouped for year,month,day.
   3.6. Interpolated the series for every 15 minutes.
   3.7. Collected the hottest and coldest day statistics and stored in csv and plotted the hottest and coldest series w.r.t 6 years.
4. **References**:
   1. For data understanding I referred to the link below.
      1.1. 1.https://www.dwd.de/EN/climate_environment/cdc/cdc_en.html
   2. Below are the code references.
      2.1. https://towardsdatascience.com/pandas-resample-tricks-you-should-know-for-manipulating-time-series-data-7e9643a7e7f3
      2.2. https://machinelearningmastery.com/resample-interpolate-time-series-data-python/
      2.3. https://pandas.pydata.org/pandas-docs/stable/user_guide/timeseries.html
      2.4. https://stackoverflow.com/questions/46011940/how-to-plot-two-pandas-time-series-on-same-plot-with-legends-and-secondary-y-axi