Savitaa Venkateswaran
301376579

REPORT

Question 2:

Part a) Produce **one or more figures** that illustrate the **max. temperature distribution over the entire globe** and enable a comparison of **different, non-overlapping time periods** (avoid overplotting).

Analysis and Description about the graph:

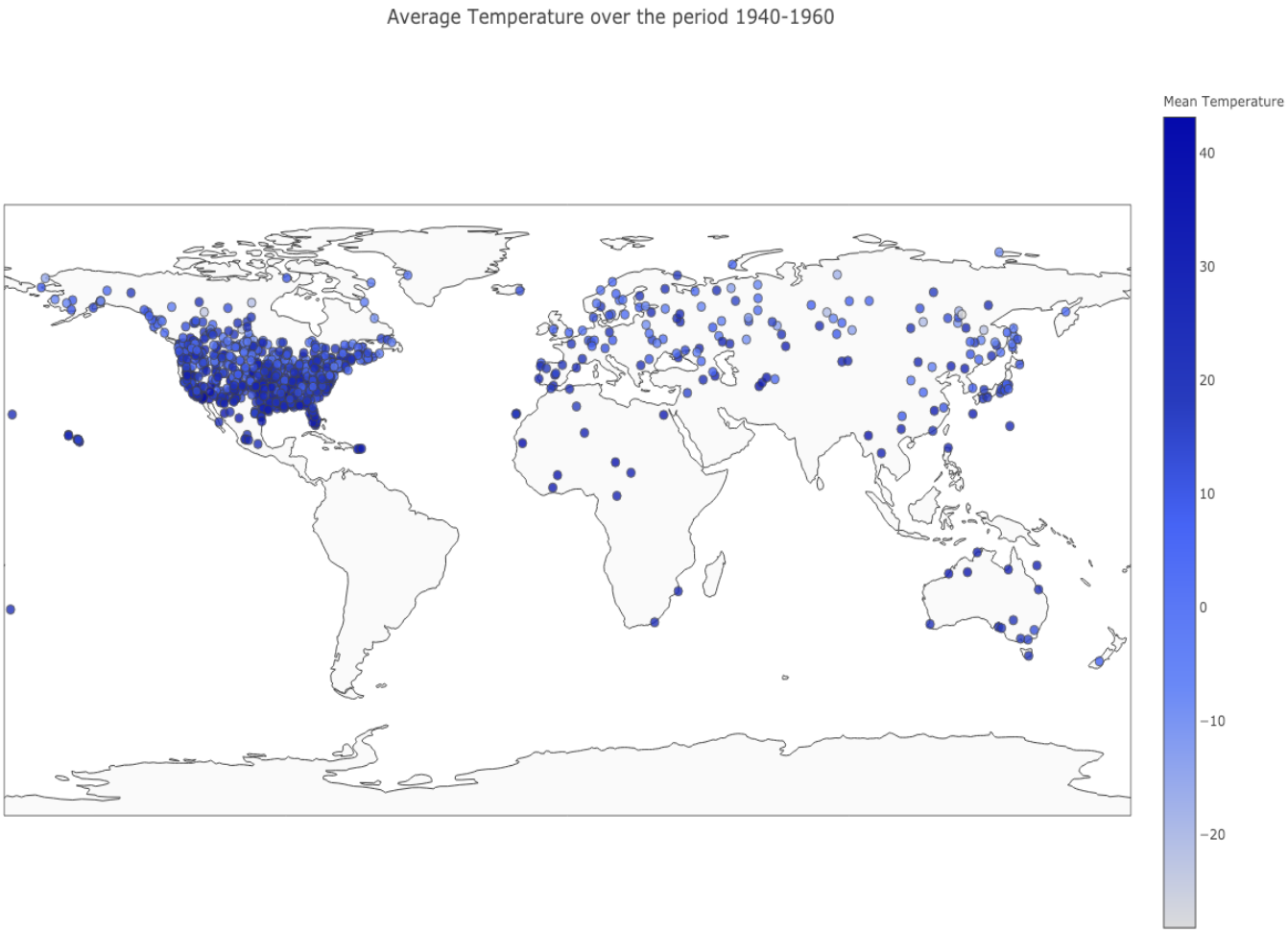

Figure 1: Average Temperature over the period of 1940-1960 using Plotly (Task A)

Figure 1 (https://plot.ly/~savitaav/22) above is a scatter plot on map using plotly over a time period (or) range of 20 years from 1940-1960. I chose this range as I wanted to contrast the

wake of 1900's to the wake of 2000's. This, I felt would be a good way to study climate change, as the difference between the two chosen time frames is substantially big (good enough).
Coming to what the plot conveys, the plot points depict places (stations) for which we had a reading of maximum temperature observed. I have taken an average/mean of the observed maximum temperature readings- whenever I found the station/place to have more than one reading (reading for maximum temperature) within the chosen time period.

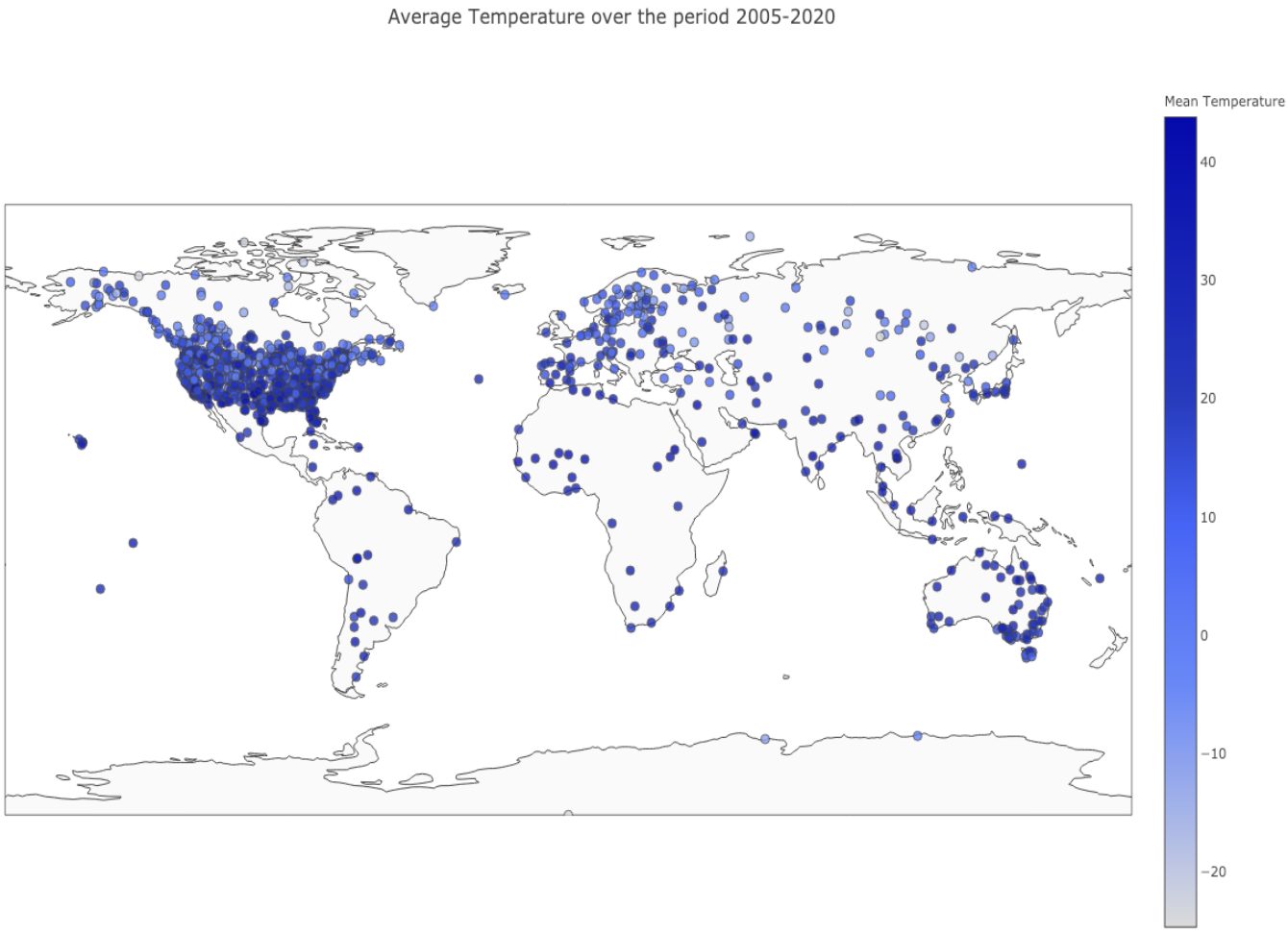

Figure 2: Average Temperature over the period of 2005-2020 using Plotly (Task A)

Figure 2 (https://plot.ly/~savitaav/16) in contrast to Figure 1, depicts the maximum temperature of places (that is for which I had readings) over a time period of 15 years from 2000-2020. I limited the range from 20 years previously to 15 years here, knowingly, as I wanted the plots to have a fairly similar number of input points to showcase. Here, as seen

(when you zoom in) for certain places there has been an increase in the average maximum temperature, whereas for certain others a slight cooling effect can be seen- if observed carefully. Also, I made sure the colorbar (legend) falls into the same range (approx -24 to 44), so that it would be easy for the audience to grab the visual information from the figure produced.

Part b) Produce two or more figures that show the result of your re-trained regression model from CMPT 732-A9, i.e. a version of the model that does not use yesterday_tmax as extra input feature:

(b1) Evaluate your model at a grid of latitude, longitude positions around the globe spanning across oceans and continents, leading to a dense plot of temperatures. This could, for instance, look something like the following. You can use a fixed day_of_year of your choice. Also, see further hints about elevation below.

Analysis and Description about the graph:

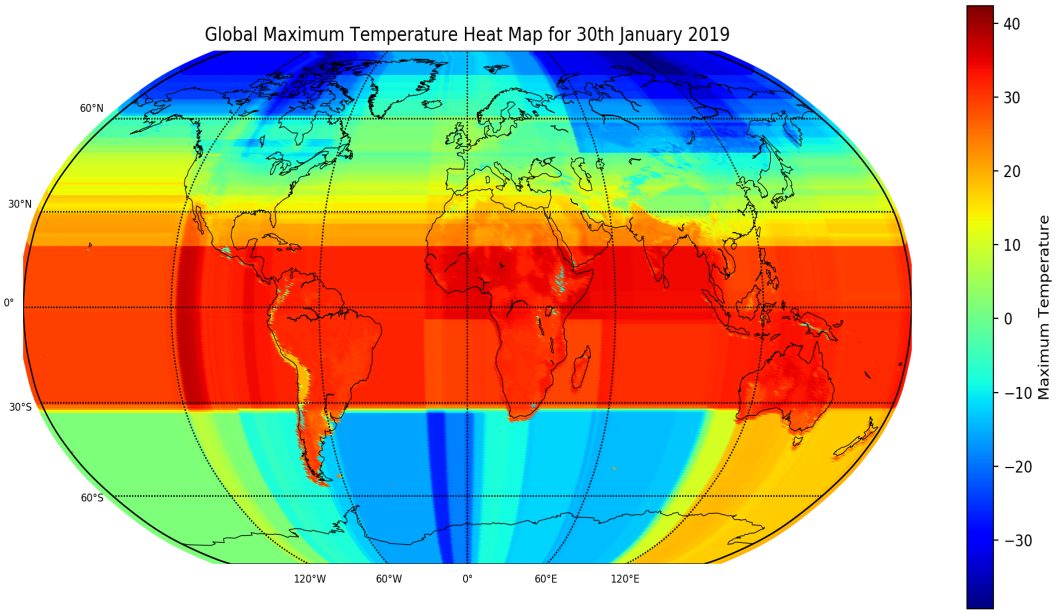

Figure 3: Global Maximum Temperature Heap Map for 30-1-2019 (Task B.1)

Savitaa Venkateswaran
301376579

Figure 3 shows the Global Heat Map produced using basemap for 30th Jan 2019 by the weather model, that has been re-trained using features except yesterday_tmax. The points produced by my model (regressor) with the help of elevation obtained for those produced points using the script provided, has been plotted as a heat map spanning across the globe. The plot depicts the maximum temperature range over the globe (spanning across oceans and land) for a chosen particular day- here, 30th jan 2019. It can been seen that the model has slight pit falls, as discrepancy can be clearly observed from the produced visualization. Overall, It seems to do a fair job, as the heat map fairly represents certain countries/continents outline and showing that the regions closer to equator are hotter than the ones closer to the poles (the zones seem to be well captured).

(b2) In a separate plot show the regression error of your model predictions against test data. In this case only use locations where data is given, i.e. you may reuse your plotting method from Part 2 (a).

Analysis and Description about the graph:

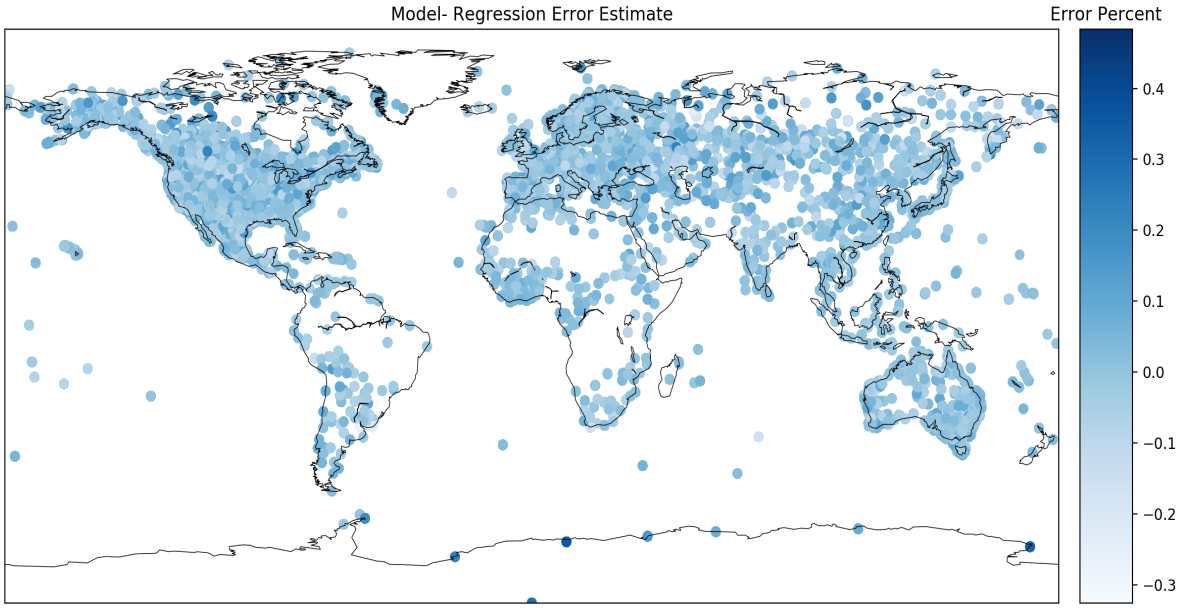

Figure 4: An estimate of the model's Regression Error (Task B.2)

Savitaa Venkateswaran
301376579

Figure 4 has been plotted using values that depict the percentage difference between predicted tmax values (by model trained) and the originally observed tmax values. The colour scale depicts the magnitude of error and points depict the places for which the error has been observed. On the whole, the error percentage in predicting the tmax value for a future date seems to be around the range of -0.2 to +0.1 percentage. So, the plus sign here indicates that the predicted value of the model is greater than the actual value and Negative sign indicates that the predicted value of the model is lesser than the originally observed tmax value (from test data). So , the model seems to be doing a decent job in predicting future values fairly well.
