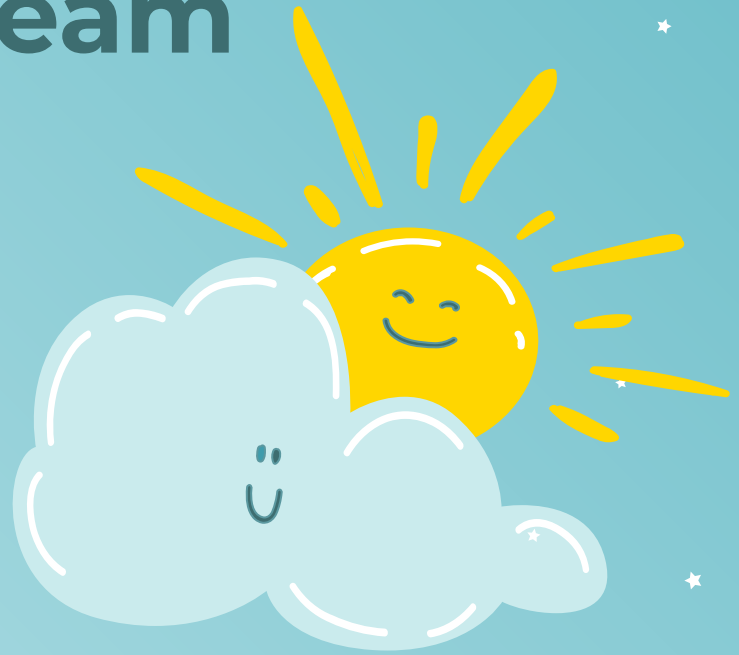
Team Awesome is Team Meteorologist

Team Awesome: Jenna Mulvihill (Leader), Hayden French, and Michael Vaden

DS 4002, 11/30/22

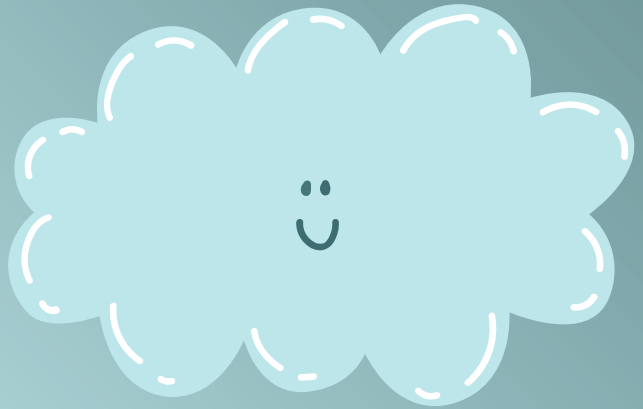


Outline

1. Motivation
2. Hypothesis
3. The Data
4. Analysis
5. Results and Conclusions
6. Next Steps
7. References

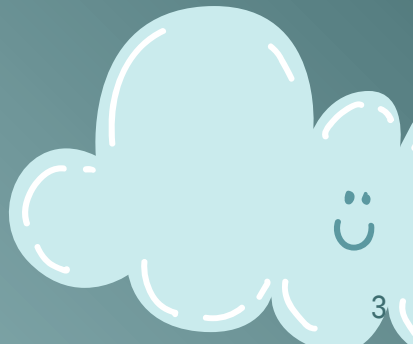
Motivation/Context

- Climate change is one of the hottest topics in politics right now
- According to the Government of Canada's website, the average temperature in Canada has increased by 1.7 degrees Celsius between 1948 and 2016 [1]
- We believe that trying to predict future temperatures for Canadian cities will help the Canadian government to know what areas are in most need of resources due to increases in temperature



Hypothesis

Using daily time series data, we can create a model to predict the average monthly temperature of Toronto, Canada from 2010-2019, with an RMSE less than 5.



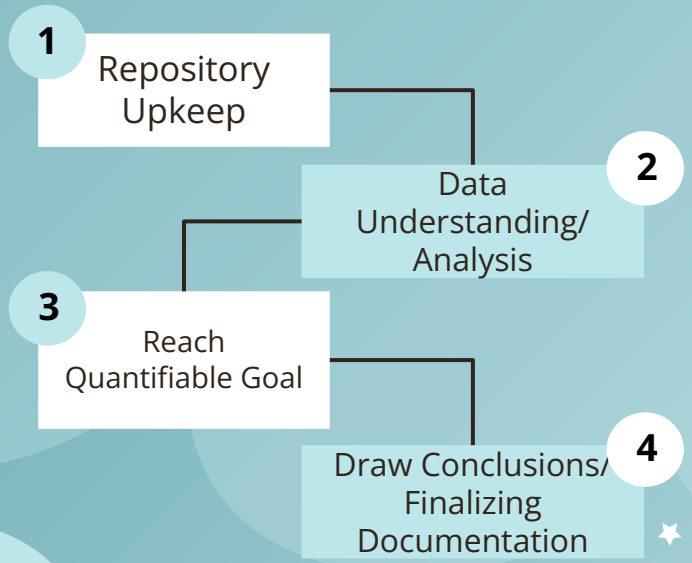
Data Acquisition/Explanation

- We found our data on Kaggle. It has 29221 rows and 9 columns, but originally had more that we subsetted from
- It is time series data
- There are temperature and precipitation variables for Vancouver, Calgary, Winnipeg, and Toronto, Canada

Variable	Description
LOCAL_DATE	Date that measurement was taken
MEAN_TEMPERATURE_[CITY]	Average temperature in Celsius in the specified city on the specified date
MEAN_PRECIPITATION_[CITY]	Average precipitation in millimeters in the specified city on the specified date

Analysis Plan/Justification

PLAN



CHALLENGES

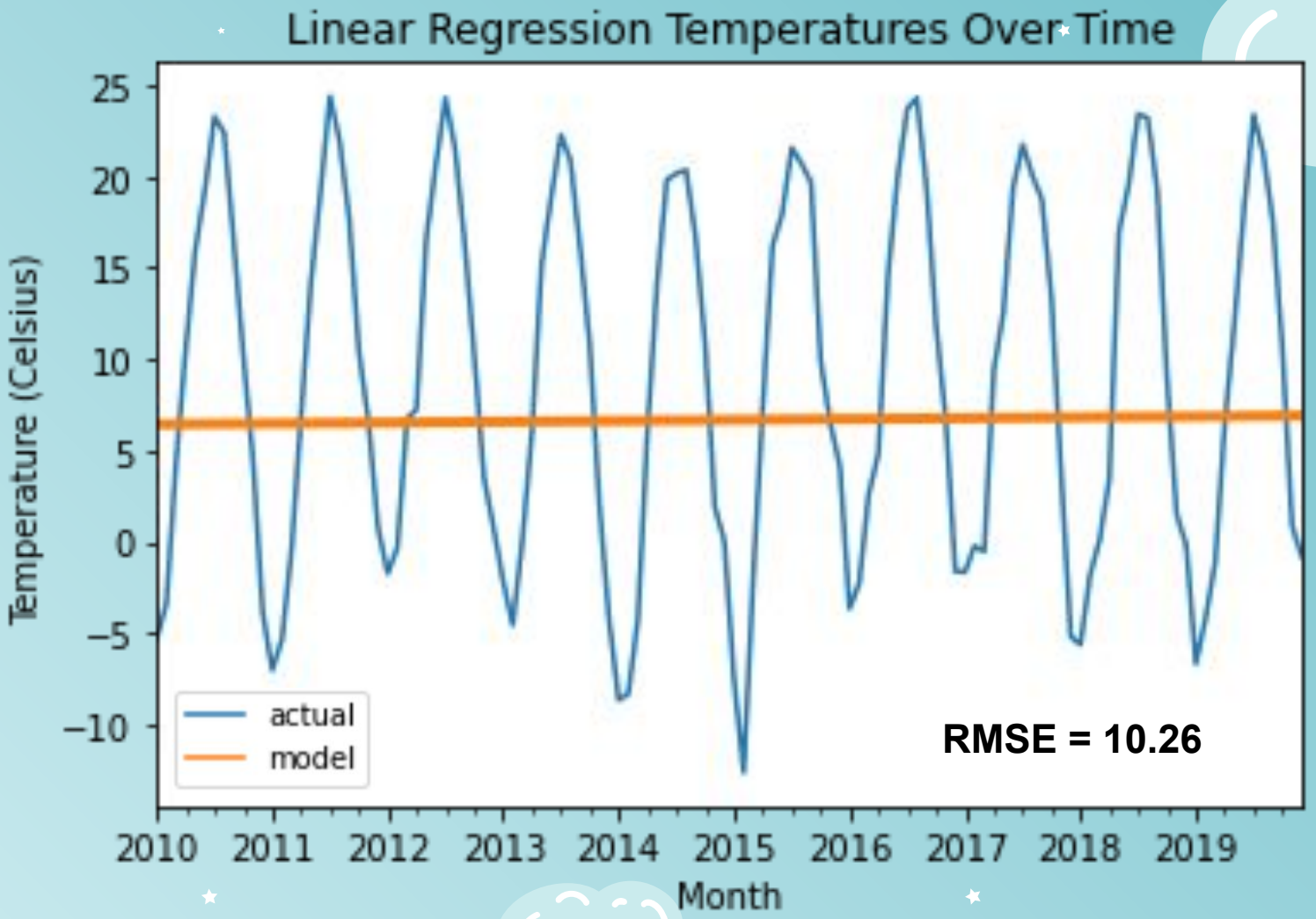
- We had overlap between different models and had to cut down the number of models we produced
- Using a SARIMA model which we have no experience with
- Comparing linear regression, random forest, and SARIMA models that are so different

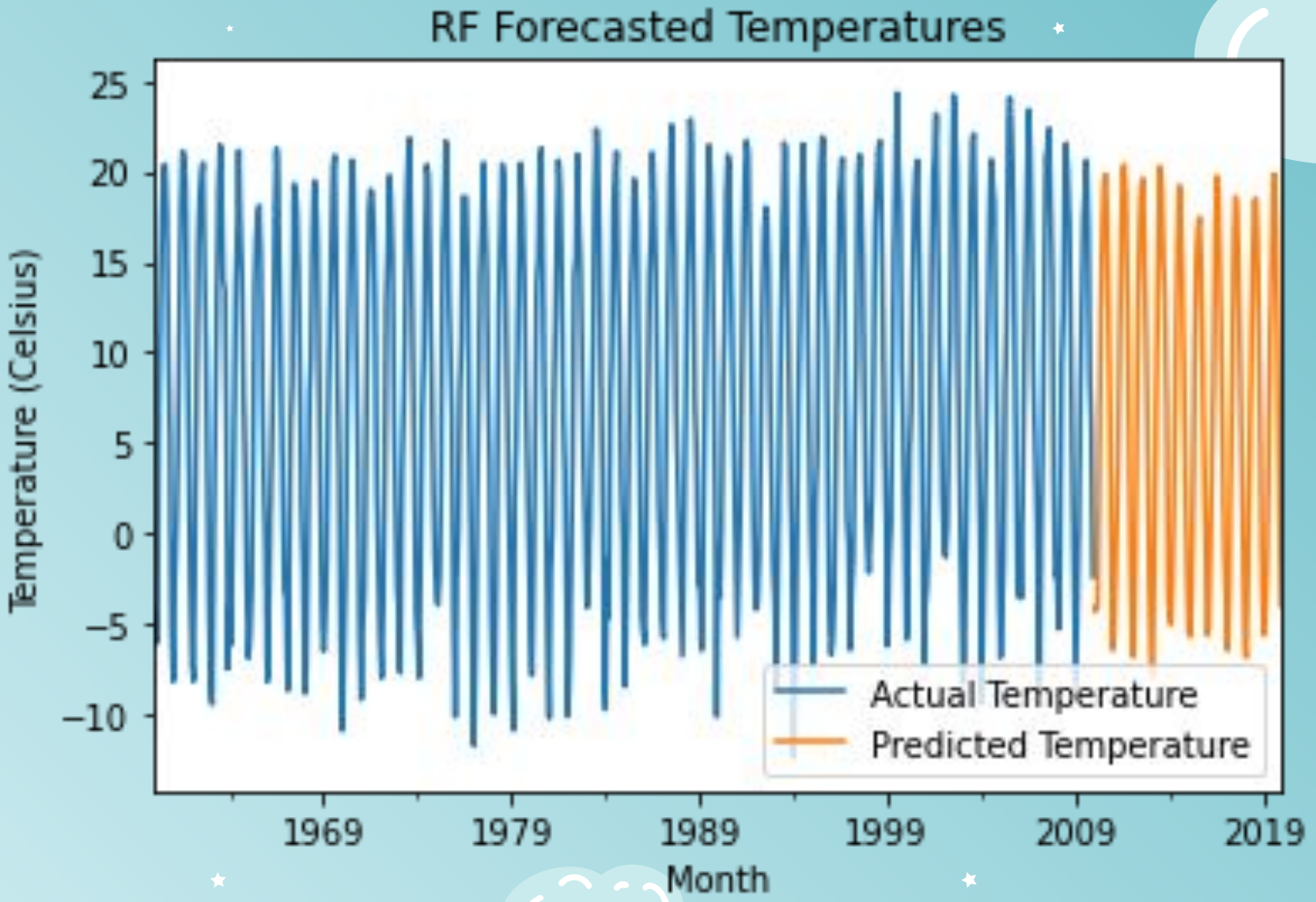
Analysis

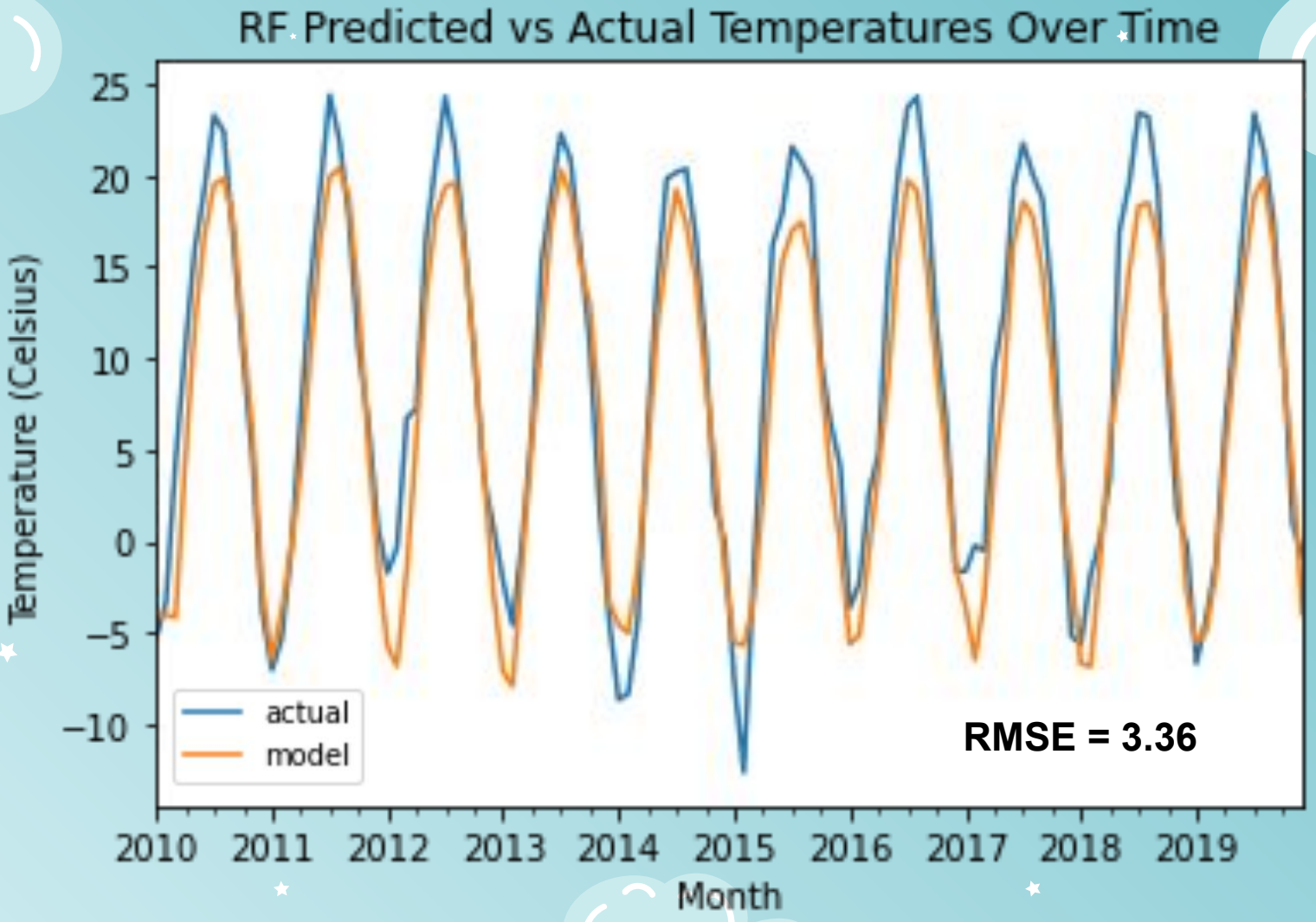
Models:

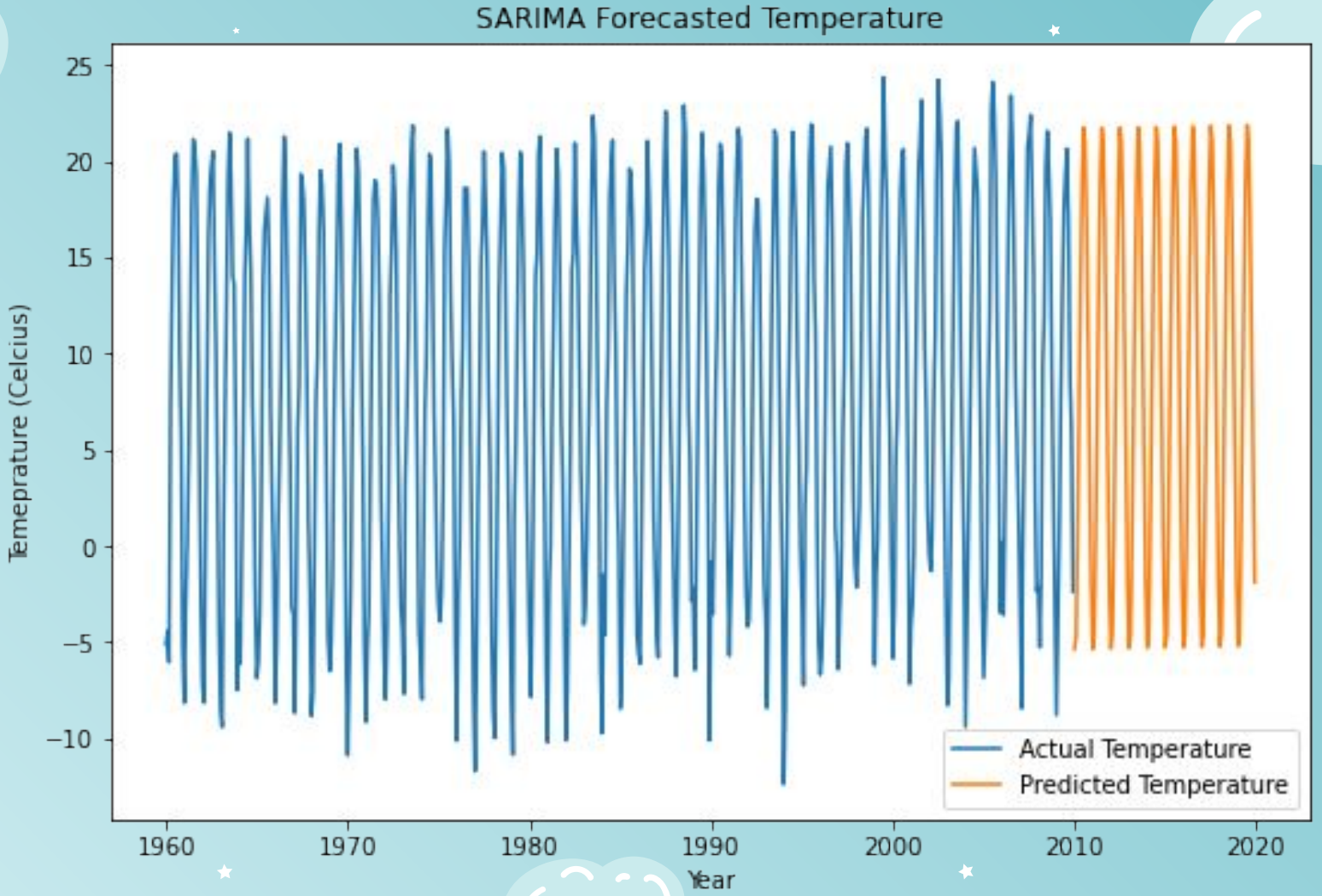
- Machine Learning Models
  - Linear Regression
  - Random Forest
- SARIMA (Seasonal Autoregressive Integrated Moving-Average)
  - A time series modeling approach which uses previous data to predict future data
  - Accounts for seasonality to best predict future values

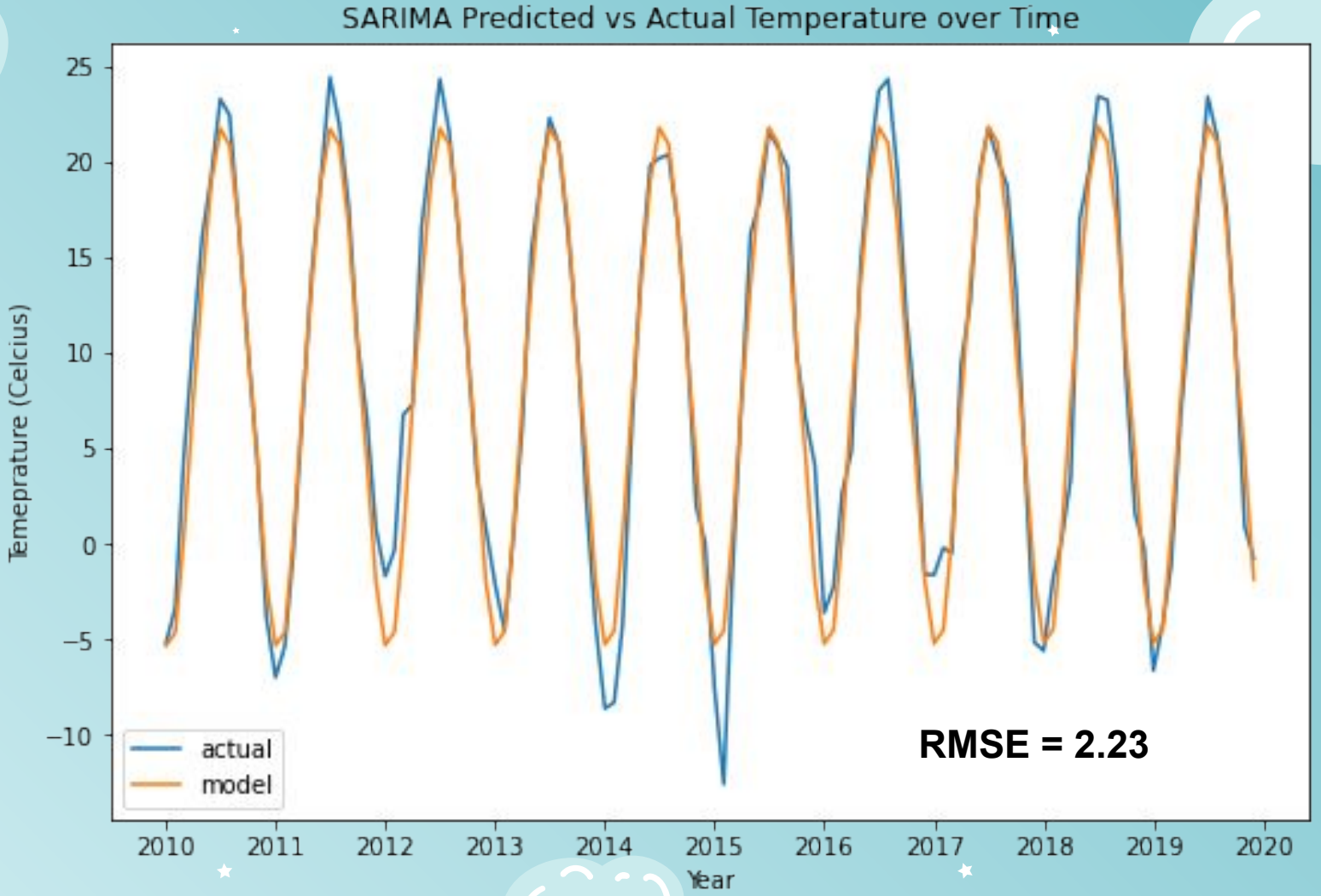
We used data from 1960-2009 to predict temperatures from 2010-2019


Linear Regression Temperatures Over Time
Temperature (Celsius)
25
20
15
10
5
0
−5
−10
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Month
actual
model
RMSE = 10.26


RF Forecasted Temperatures
Temperature (Celsius)
25
20
15
10
5
0
−5
−10
1969
1979
1989
1999
2009
2019
Month
Actual Temperature
Predicted Temperature


RF Predicted vs Actual Temperatures Over Time
Temperature (Celsius)
25
20
15
10
5
0
−5
−10
actual
model
RMSE = 3.36
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Month


SARIMA Forecasted Temperature
Temeprature (Celcius)
25
20
15
10
5
0
−5
−10
1960
1970
1980
1990
2000
2010
2020
Year
Actual Temperature
Predicted Temperature


SARIMA Predicted vs Actual Temperature over Time
Temeprature (Celcius)
25
20
15
10
5
0
−5
−10
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Year
actual
model
RMSE = 2.23

Potential Biases/Caveats

- The SARIMA model predicted for all of the temperatures to be the same for the ten years, which we think could be an issue as temperatures could change over the years
- RMSE isn't always the tool used to find error in SARIMA models, but we used it here to be able to compare to our other models
    - MAPE is the standard, but didn't work well because our data is close to zero and has both positive and negative values

Results/Conclusions

Models:

- Linear Regression: RMSE of 10.26
- Random Forest: RMSE of 3.36
- SARIMA: RMSE of 2.23

- Our SARIMA model was the model that worked the best, which we expected as it accounts of seasonality
- We are able to accept our hypothesis!

Next Steps

- Test out our model on other cities in Canada to see how accurate it can predict temperatures
- See if there may be a model that is able to predict temperatures better than our model did
- New questions:
  - Can we use this same model for in the United States or other countries?
  - Can we use a model similar to this for predicting precipitation?

References/Acknowledgements

[1] E. and C. C. Canada, “Government of Canada,” Canada.ca, 07-Nov-2022. [Online]. Available: https://www.canada.ca/en/environment-climate-change.html. [Accessed: 09-Nov-2022].
[2] E. and C. C. Canada, “Government of Canada,” Canada.ca, 09-Apr-2019. [Online]. Available: https://www.canada.ca/en/environment-climate-change/services/climate-change/canadian-centre-climate-services/basics/trends-projections/changes-temperature.html. [Accessed: 02-Nov-2022].
[3] “Time Series Forecasting methods,” InfluxData, 19-Sep-2022. [Online]. Available: https://www.influxdata.com/time-series-forecasting-methods/. [Accessed: 09-Nov-2022].
[4] “Different types of time series decomposition.” [Online]. Available: https://towardsdatascience.com/different-types-of-time-series-decomposition-396c09f92693. [Accessed: 09-Nov-2022].
[5] “Time Series Forecast Error Metrics you should know.” [Online]. Available: https://towardsdatascience.com/time-series-forecast-error-metrics-you-should-know-cc88b8c67f27. [Accessed: 09-Nov-2022].

We would like to thank Professor Alonzi and Harsh for their assistance on this project