South East Brazil - Predicting Future Temperature

Calvin Yu

Problem/Need

- Government of Brazil wants to keep track of the weather for its territory
- Some of the data are missing, the government wants me to fix it and would gladly want me to forecast the future temperature

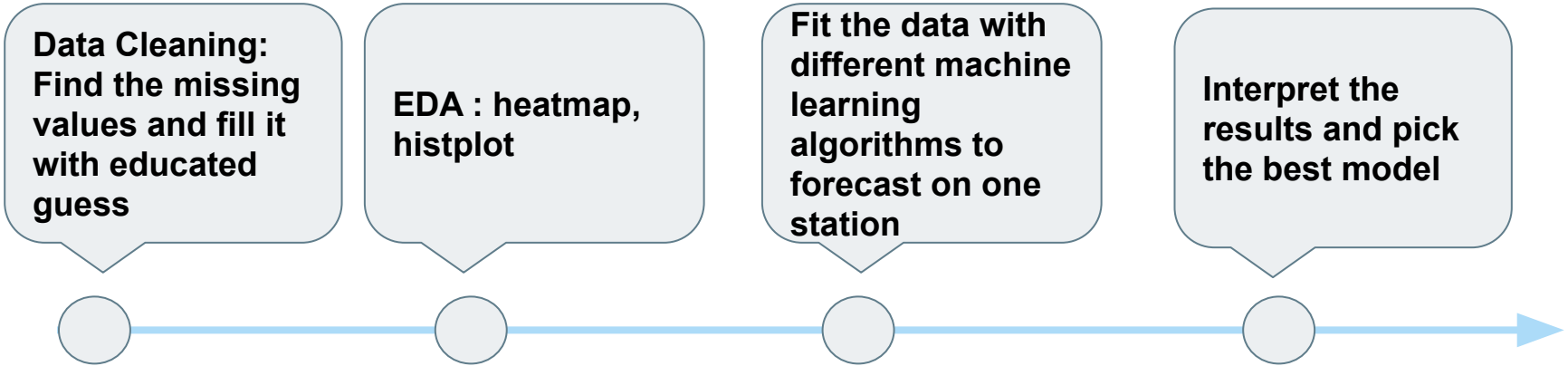
Impact Hypothesis

- Fixing the missing data can give Brazil Government a better outlook of how the weather was like in the past
- Fixing the missing data can also help us to create a better model to forecast the future

Data

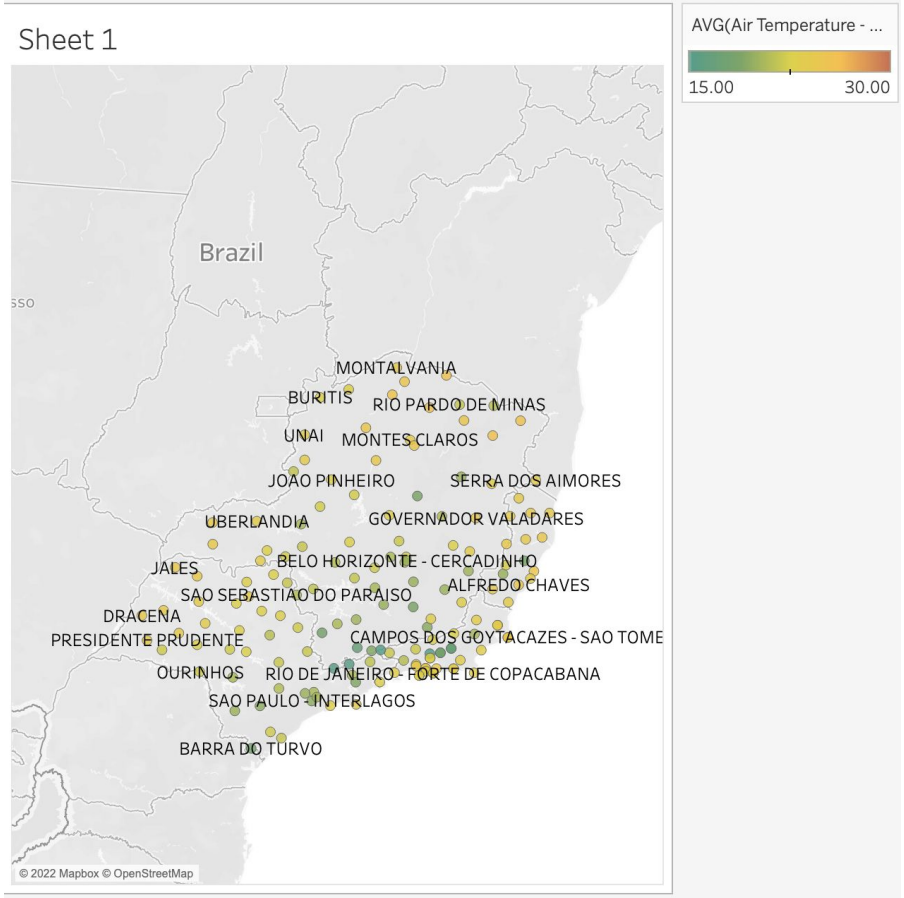
- Southeast Brazil weather data
- 15345216 rows with 27 features and 1 target variables
- From 2000 to 2021

Solution Path



Algorithms

- Linear Regression
- Ridge Regression
- Lasso Regression
- ARIMA Time-Series algorithms



Results

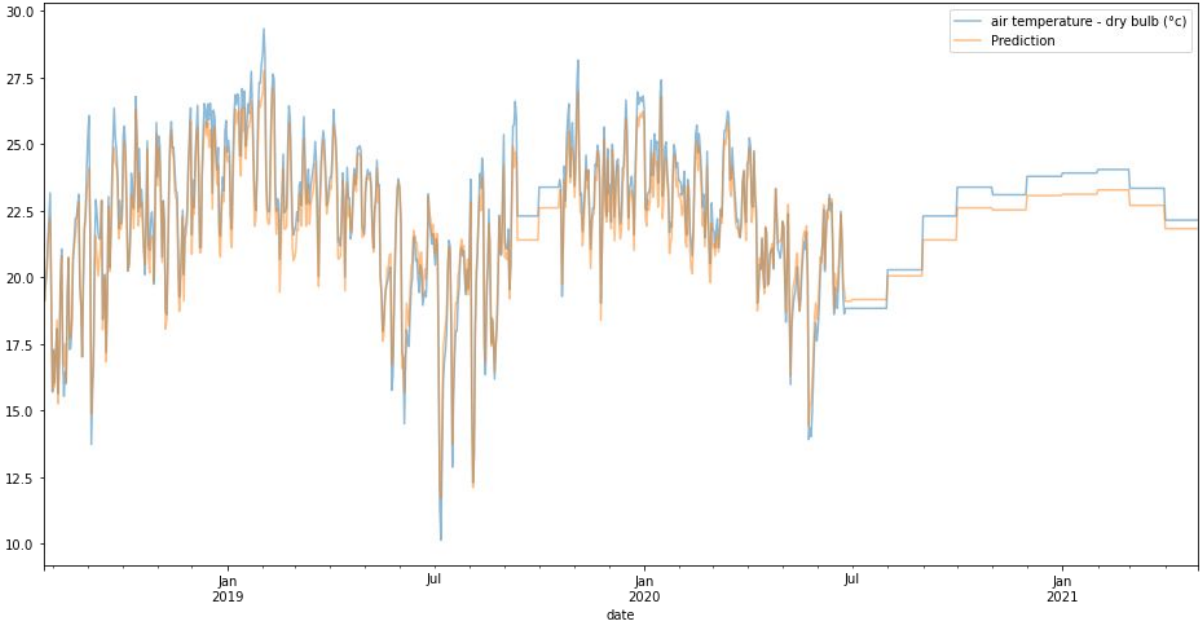
```
Lasso  0.9464285657716874
Ridge  0.9539270926051644
Linear Regression  0.9527911659207182
SARIMAX time series 0.9392093416079489
```

- Ridge is the best model , but other models are relatively accurate too

Results con.

- Linear regression , Ridge, and Lasso are regression algorithms that doesn’t take time variable and predict the results based on existing features
- Time series takes time variable and forecast it with exogenous features

Results con.


air temperature - dry bulb (°c)
Prediction
30.0
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
Jan
2019
Jul
Jan
2020
Jul
Jan
2021
date

Further Step

- Try to forecast more stations

- Try to take latitude and longitude as considering to forecast the whole dataset