

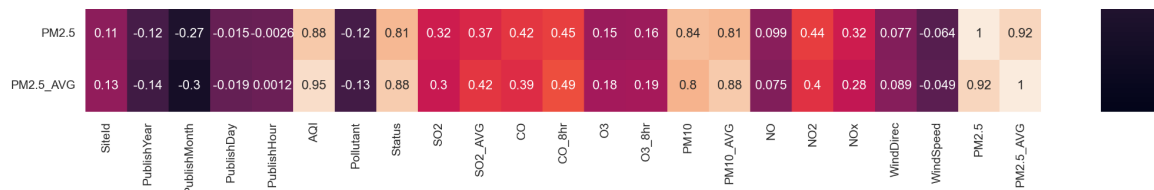
BDA - Final Project

PM2.5 Prediction - Group 24

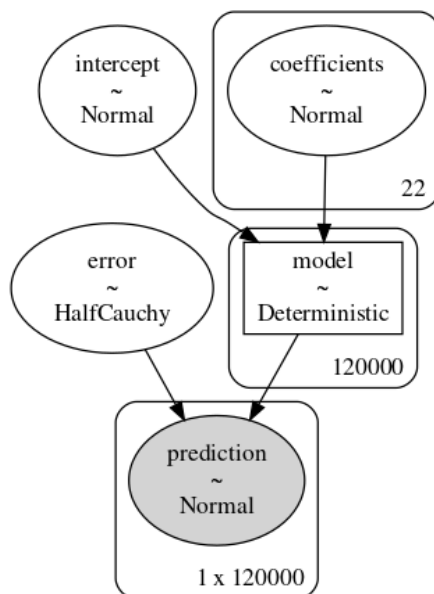
Concept

From the heatmap we can observe that many variables have a high linear correlation coefficient to the PM2.5, so we decide to use GLM.

heatmap:

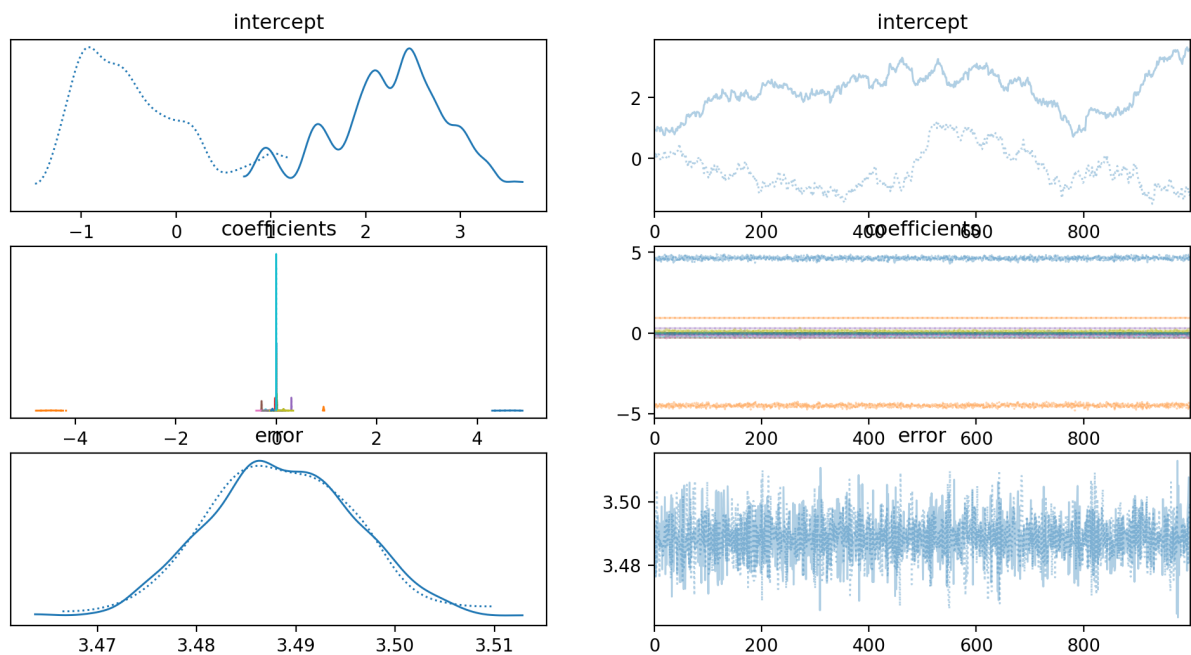


the concept of our model:



Performance

Kruschke diagram:



Model:

$$y = 0.926 + [0.001 \ 0 \ 0.001 \ 0.001 \ -0.02 \ 0.013 \ -0.12 \ -0.024 \ 0.145 \ -0.21 \ 4.64 \ -4.48 \\ 0.009 \ -0.019 \ 0.302 \ -0.292 \ -0.15 \ -0.023 \ 0.089 \ -0.001 \ -0.075 \ 0.942] * x$$

MSE:

```
def mean_square_error(target_prediction: list, target_truth: list) -> float:
    return np.square(np.subtract(target_prediction, target_truth)).mean()
```

```
print('MSE: {:.2f}'.format(mean_square_error(predictions, test_target)))
```

MSE: 12.63