

Practice Problem Statements

1. Climate changes day by day, year to year. One way to understand this is by analysis and understanding the heat index of an area. Temperature data is a small part of it. But, The findings can lead to bigger and more serious inventions and outcomes!

From this dataset , it would be interesting to find out:

What was the max and min temperature?

How was the outside temperature related to inside temperature? any relation between the two?

What was the variance of temperature for inside and outside room temperature?

What is the trend in the data?

Can you use Time Series Forecast algo to predict the next scenario?

Which was the hottest/coolest month ?

Any warning signals from climate disaster ?

As well as many others...

Dataset Description:

The following dataset contains the temperature readings from IOT devices installed outside and inside of an anonymous Room (say - admin room). The device was in the alpha testing phase. So, It was uninstalled or shut off several times during the entire reading period (28-07-2018 to 08-12-2018). This random interval recordings and few mis-readings (outliers) makes it more challenging to perform analysis on this data.

Technical Details:

columns = 5 | Rows = 97605

id : unique IDs for each reading

room_id/id : room id in which device was installed (inside and/or outside) -> currently 'admin room' only for example purpose.

noted_date : date and time of reading

temp : temperature readings

out/in : whether reading was taken from a device installed inside or outside of the room?

Link to the Dataset:

<https://drive.google.com/file/d/1u50Hf9ZtNIBKbn8Zghb3ZjfpEObUoDtP/view?usp=sharing>

2. Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. This dataset describes the listing activity and metrics in NYC, NY for 2019.

Content

This data file includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions.

Acknowledgements

This public dataset is part of Airbnb, and the original source can be found on this website.

Inspiration

What can we learn about different hosts and areas?

What can we learn from predictions? (ex: locations, prices, reviews, etc)

Which hosts are the busiest and why?

Is there any noticeable difference of traffic among different areas and what could be the reason for it?

Use the following dataset to do your first hand analysis.

<https://drive.google.com/file/d/1iloTrxN6Z2ZoWntyNwumycwvglWHholw/view?usp=sharing>

3. Build a simple linear regression model to predict the price of a house from the dataset given below. Also find the goodness of fit of the model that you have fitted on the data.

Outline

Preprocessing

Understanding the problem and the data available

Normality and skewness

Missing values

Dummy encoding

Rescaling

Regression

Linear regression
L1 regularization
L2 regularization
ElasticNet regularization

Dataset:

<https://drive.google.com/file/d/1Ld6BQsxiB8mNCgm5kO09cCIUTXmZK-hc/view?usp=sharing>