

# FORMULA AI HACKATHON '22

Ishneet Dhiman  
Vishal Dhiman  
Vudit Agrawal

# TOPICS COVERED

OBJECTIVE

DATASET

EDA

MODEL

CONCLUSION

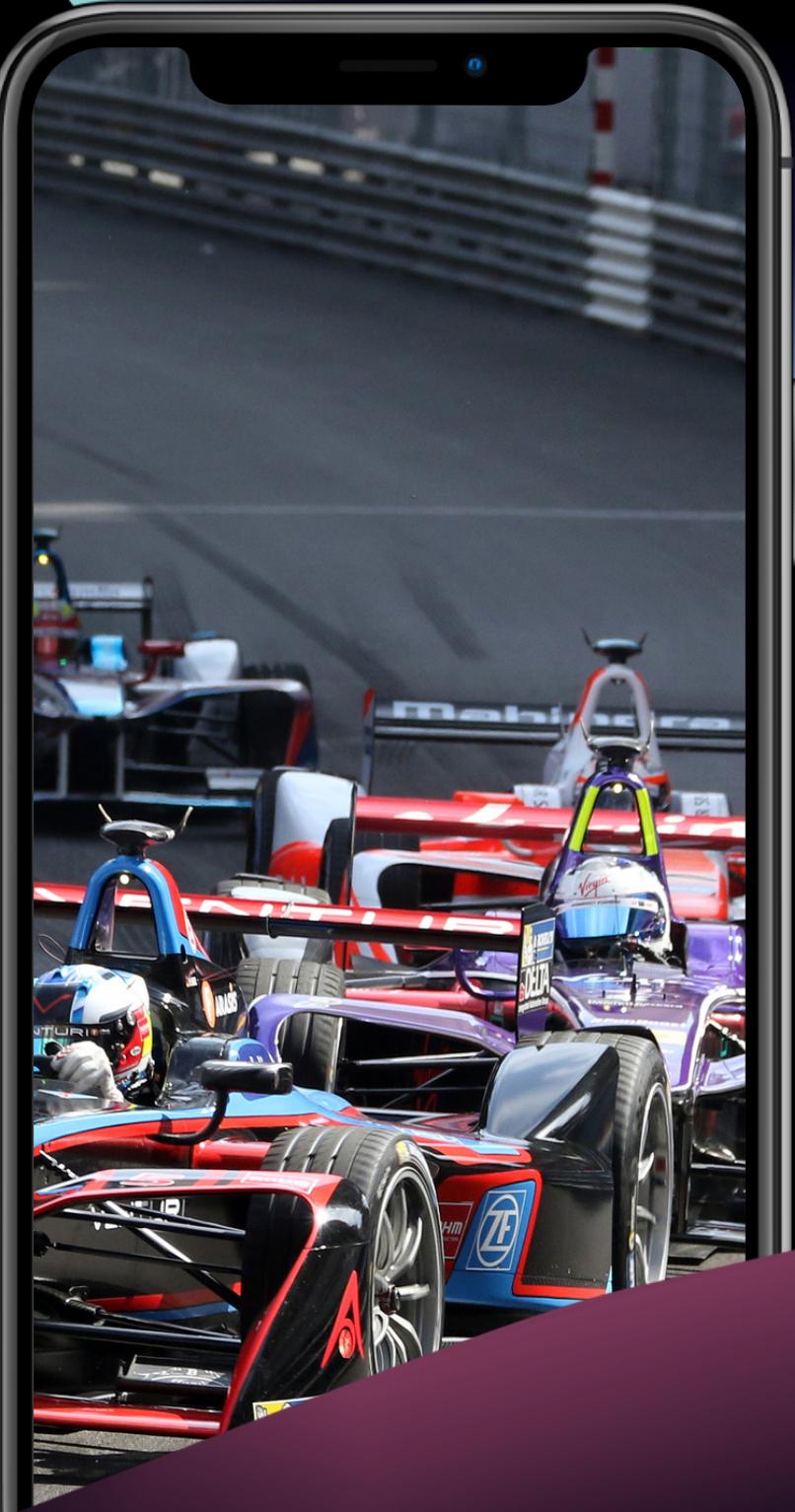
# AGENDA

# OBJECTIVE

[Go Back to Agenda Page](#)

# WEATHER AND RAIN OCCURENCE PREDICTION

Using encoded LSTM post data pre-processing and exploratory data analysis to get weather and rain probability prediction



# DATASET

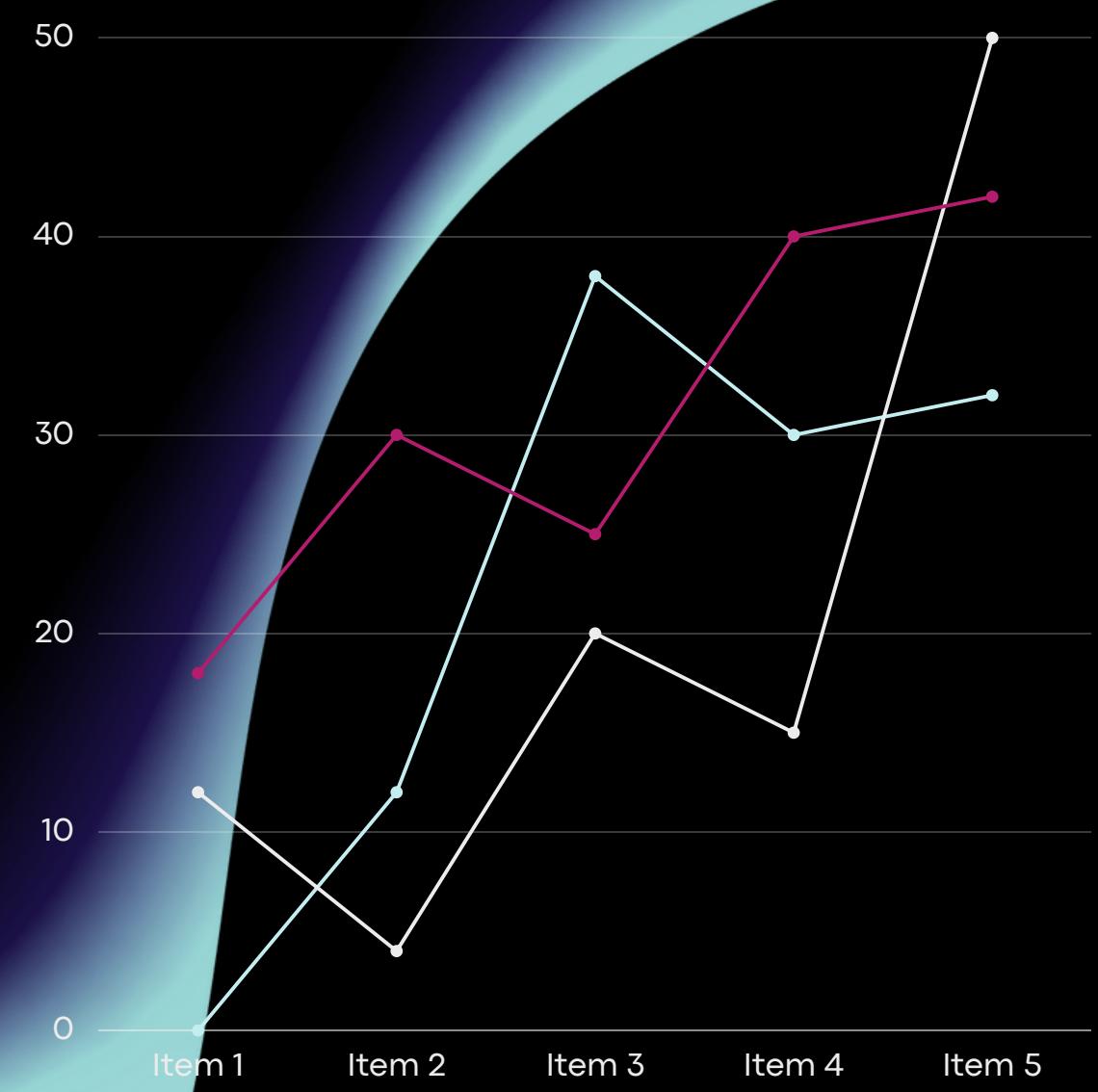
[Go Back to Agenda Page](#)

The dataset contains data pertaining to factors affecting track, weather and rainfall for different tracks, zones and times so that an inclusive prediction can be made.

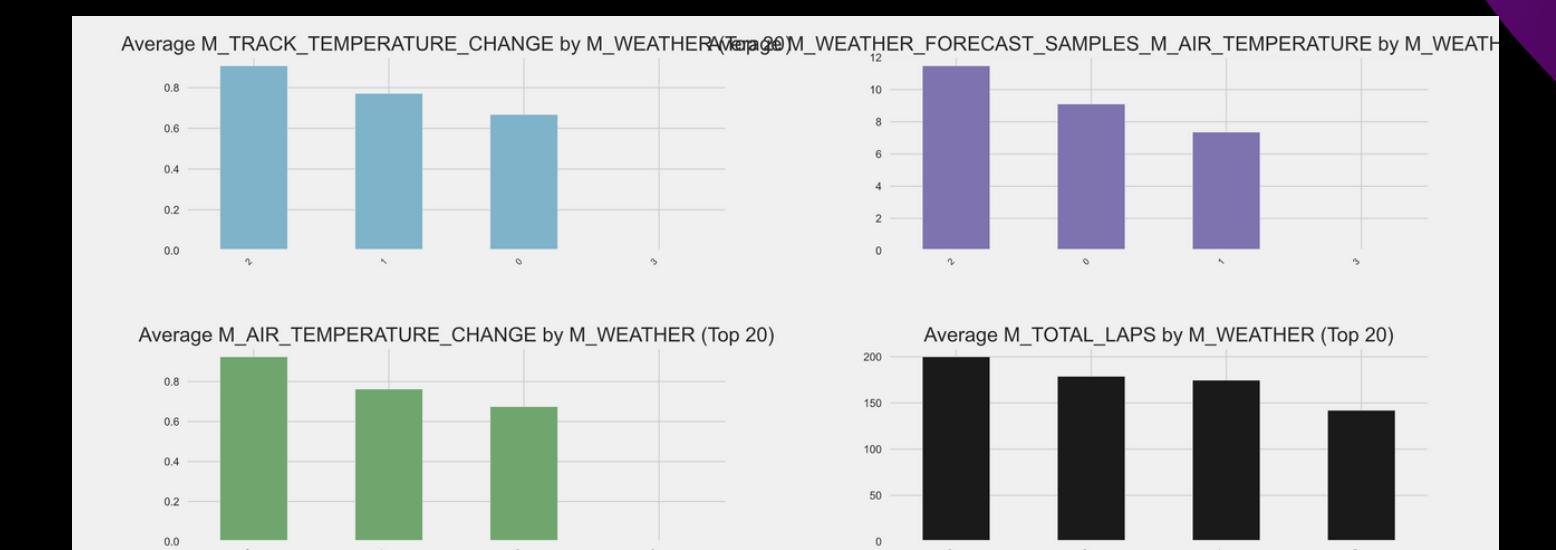
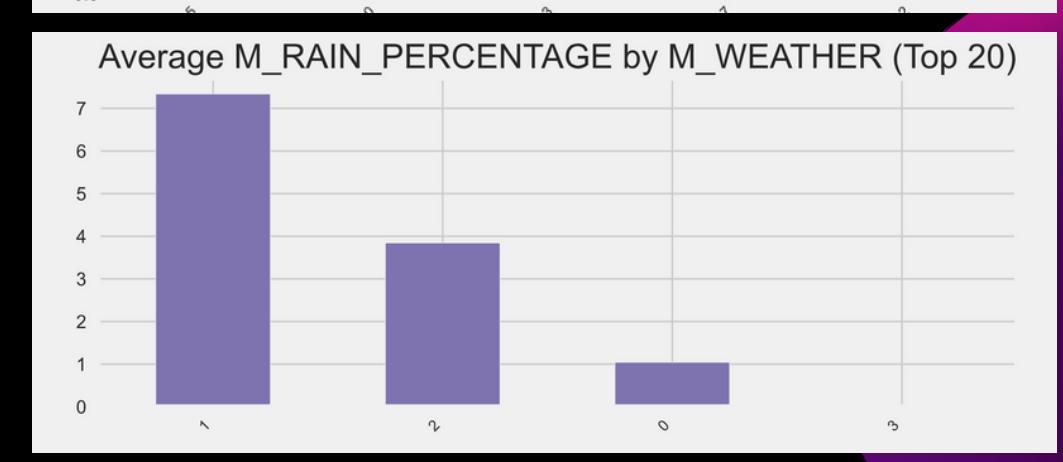
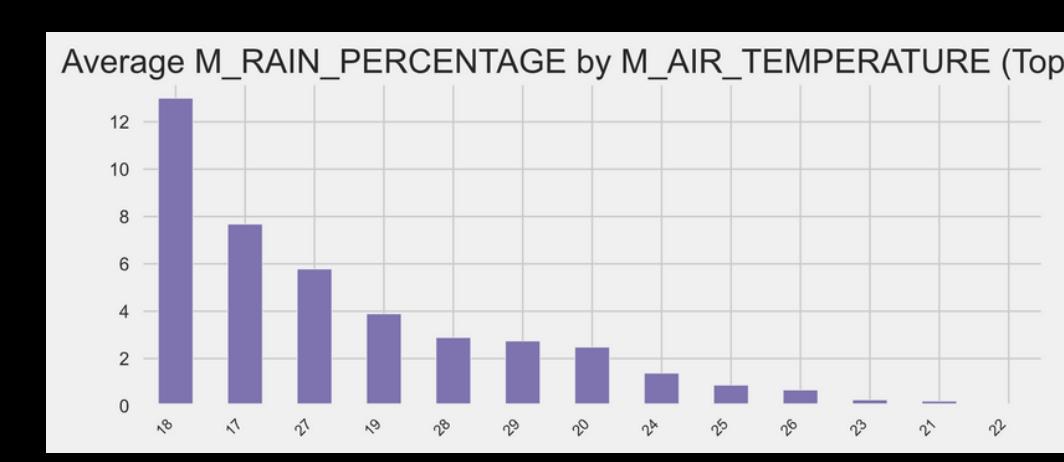
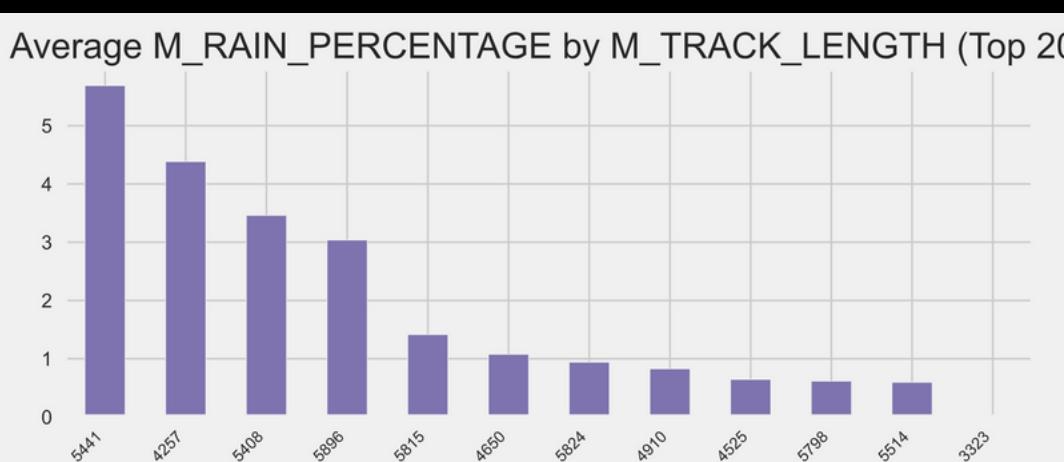
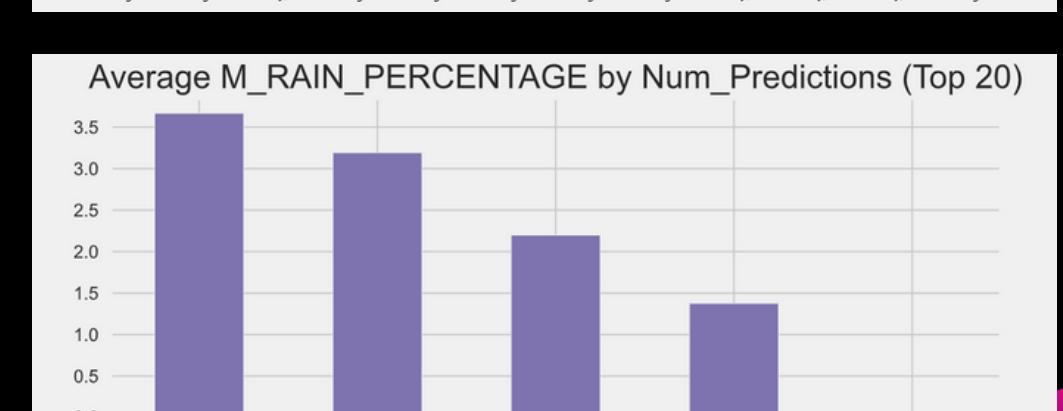
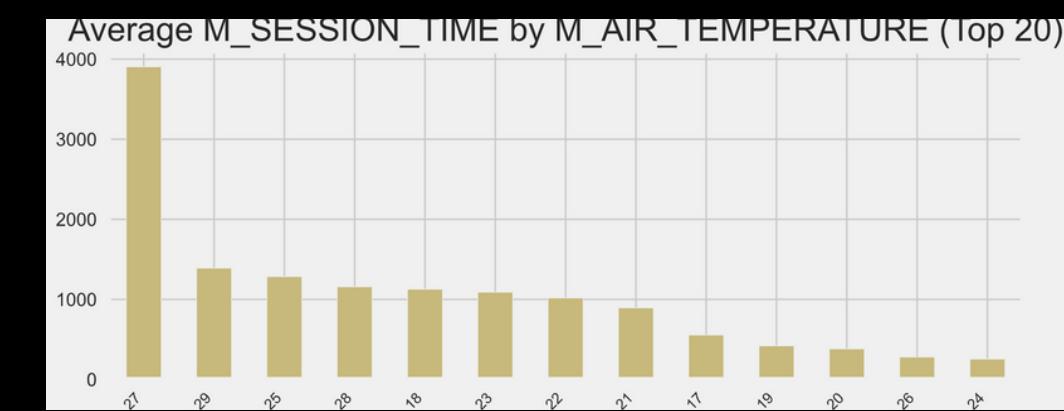
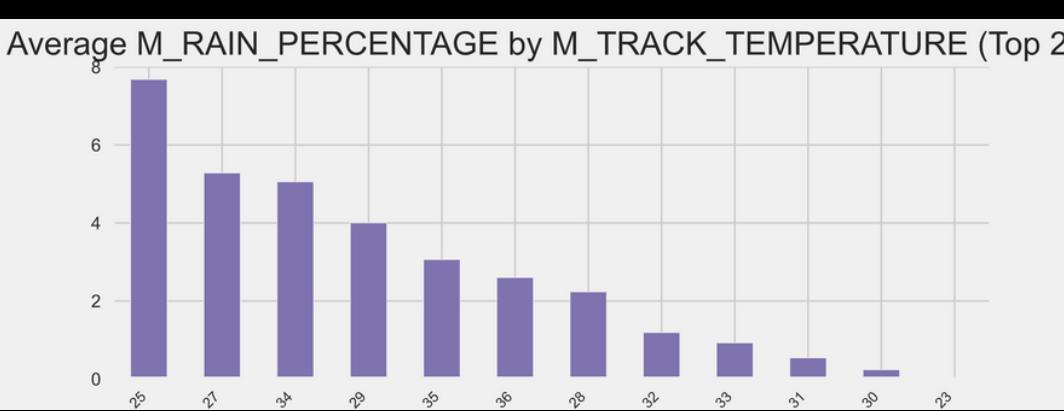
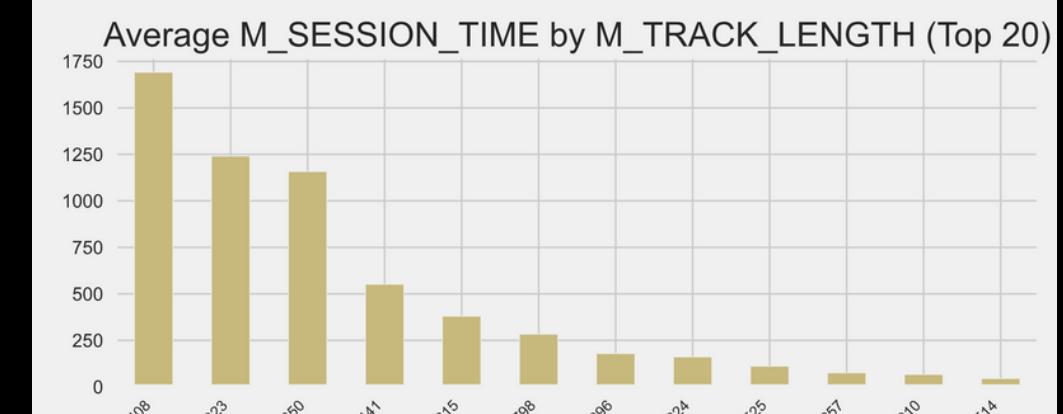
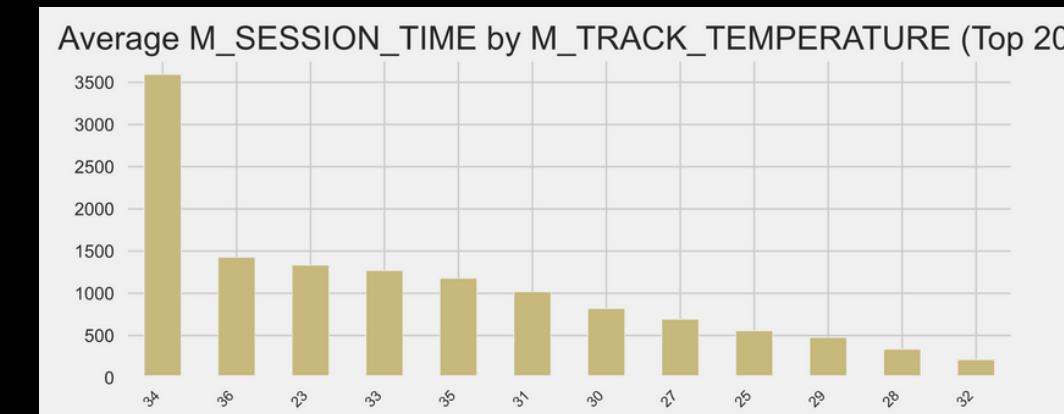
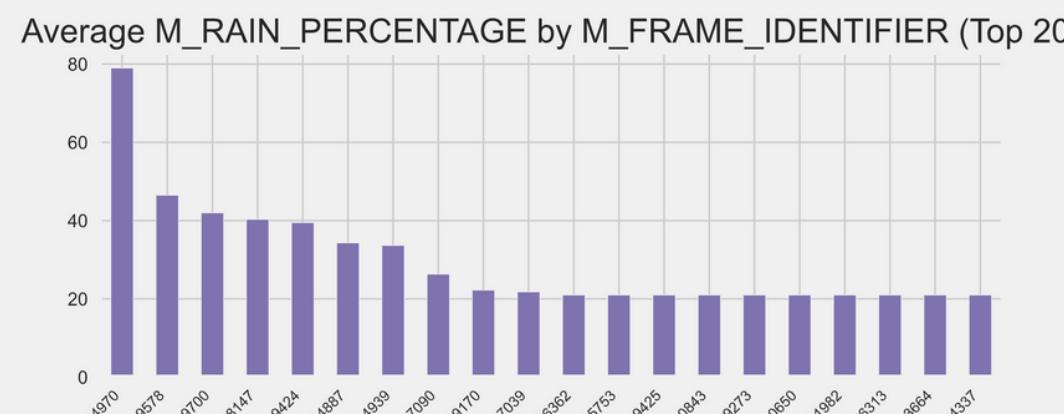
- It contains over 2.5 million rows and over 55 columns.
- Numerous redundant data are data with high collinearity which was removed.
- Data with very high zeros were also removed.
- Some features containing few number of zeroes were filled in by interpolation of present data.
- Outliers were taken care of by bell-curve principle and hence sample processed data was obtained.

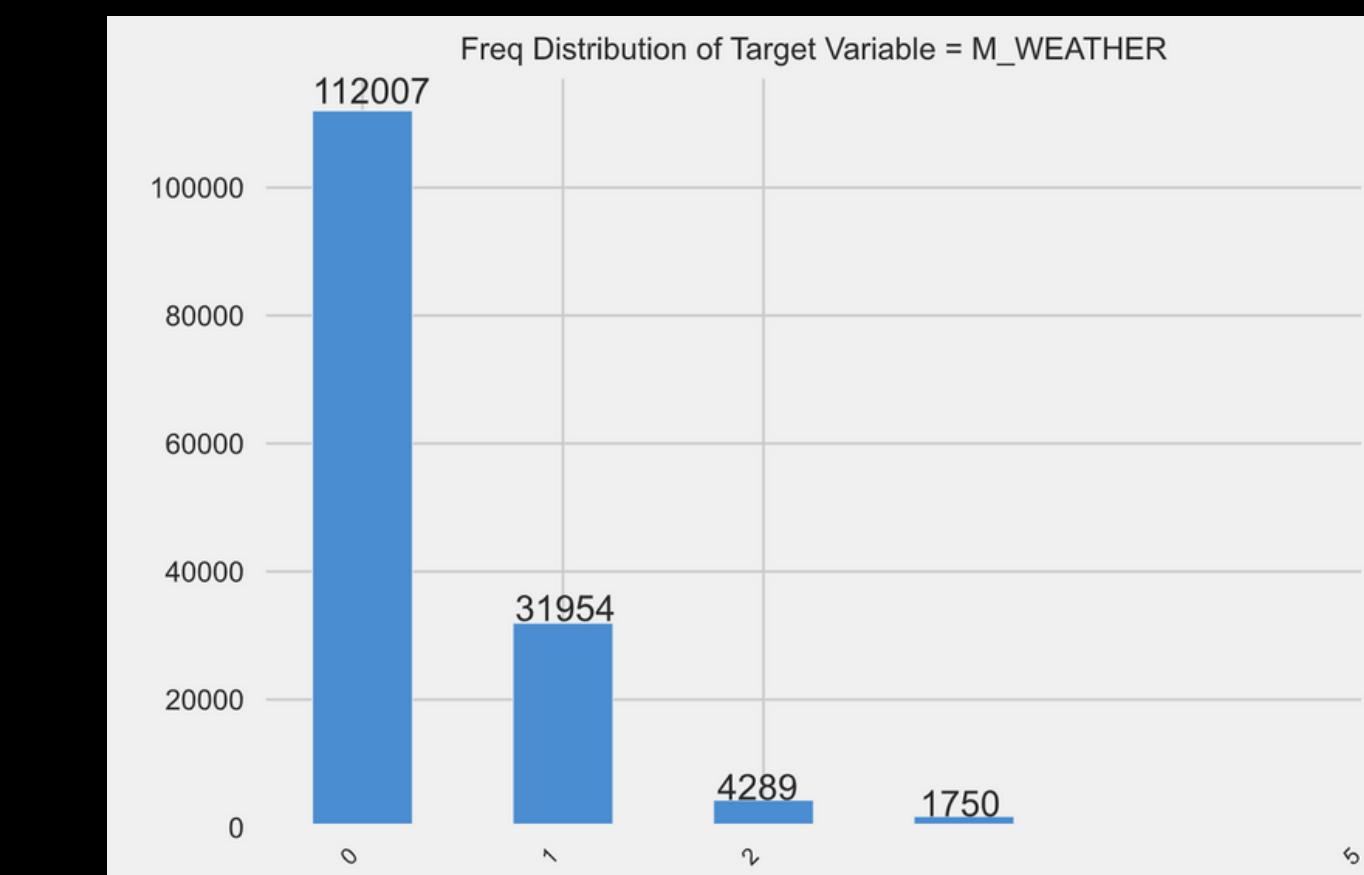
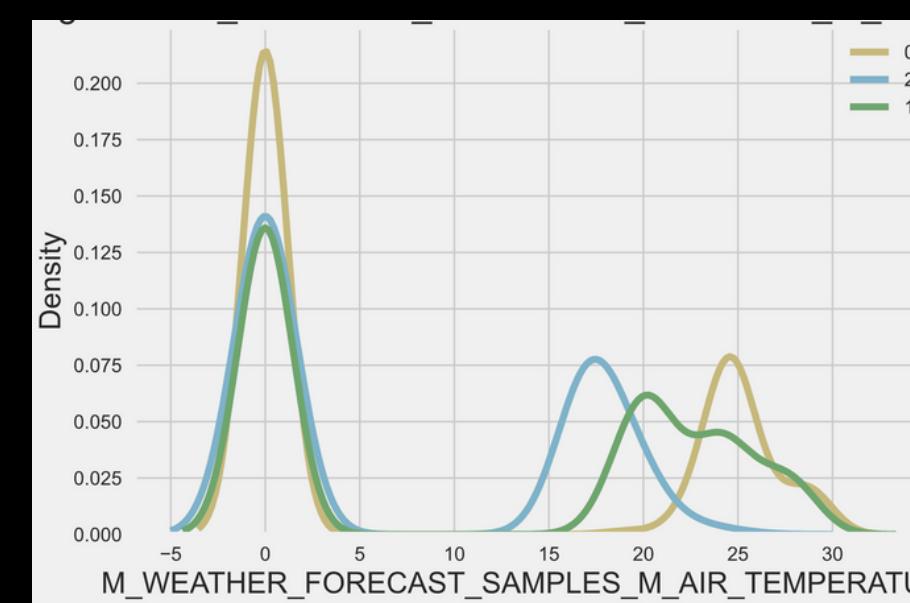
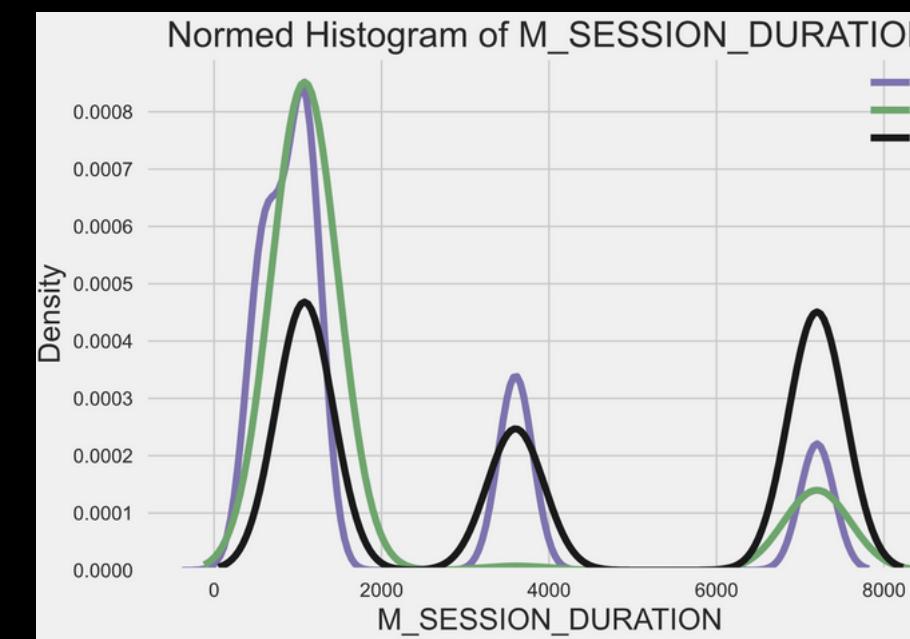
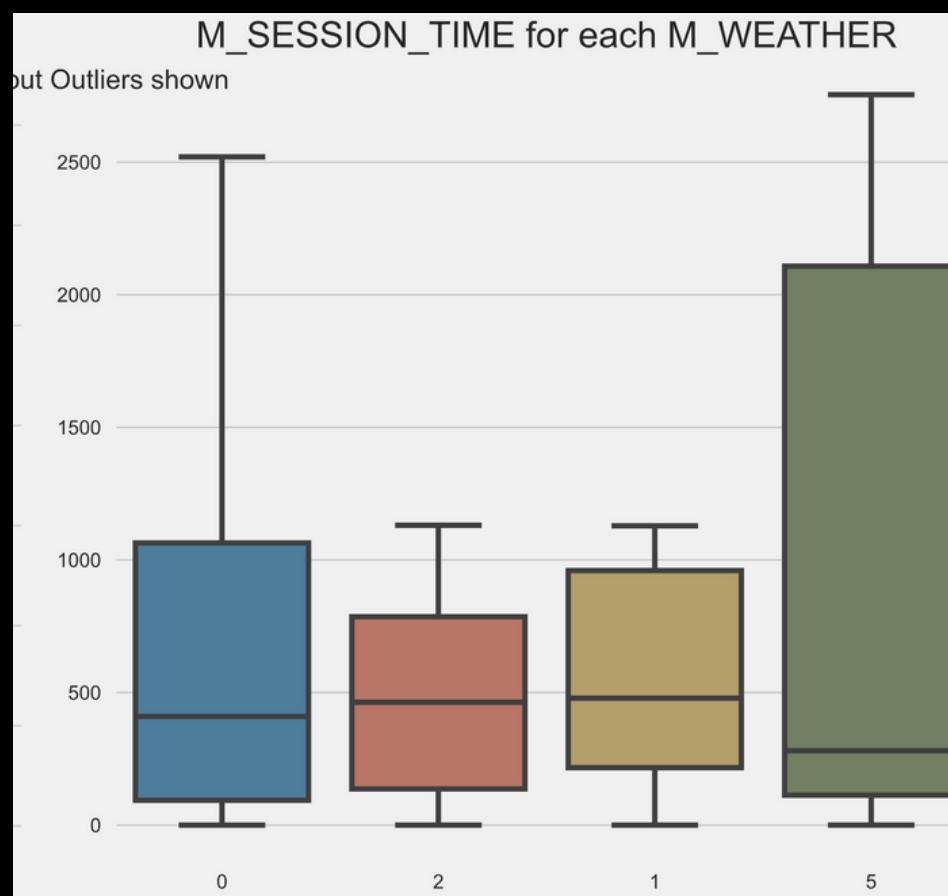
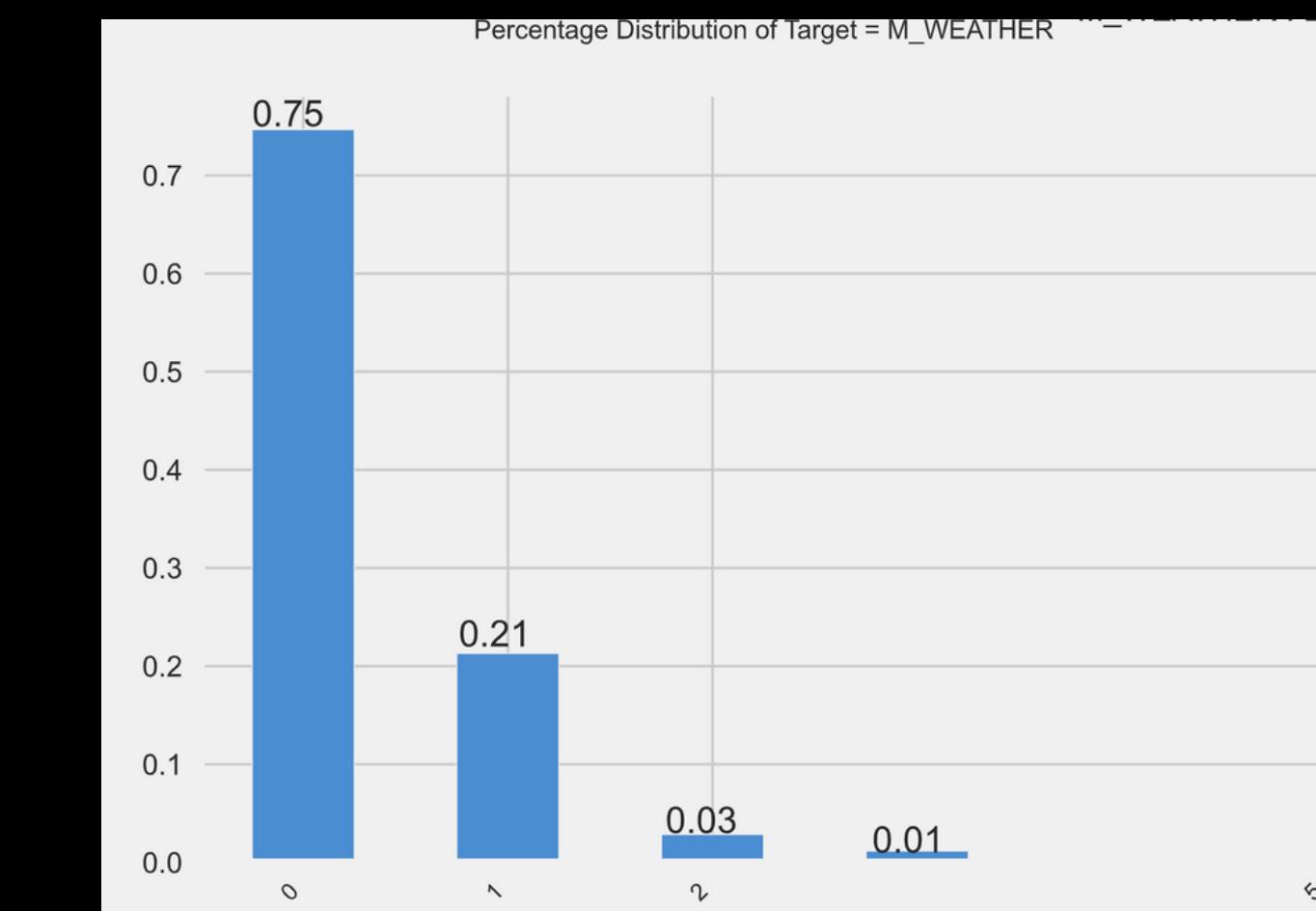
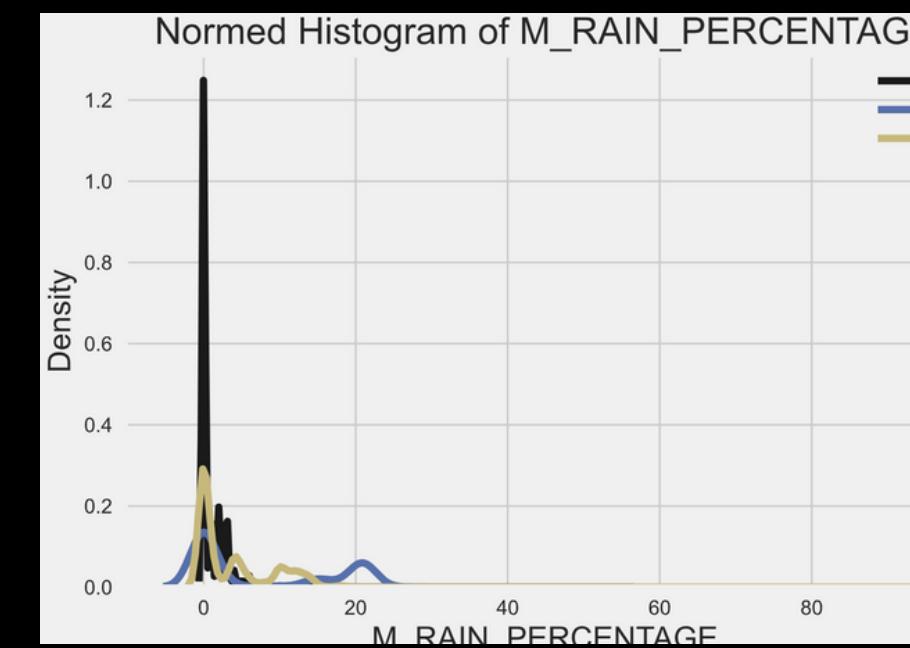
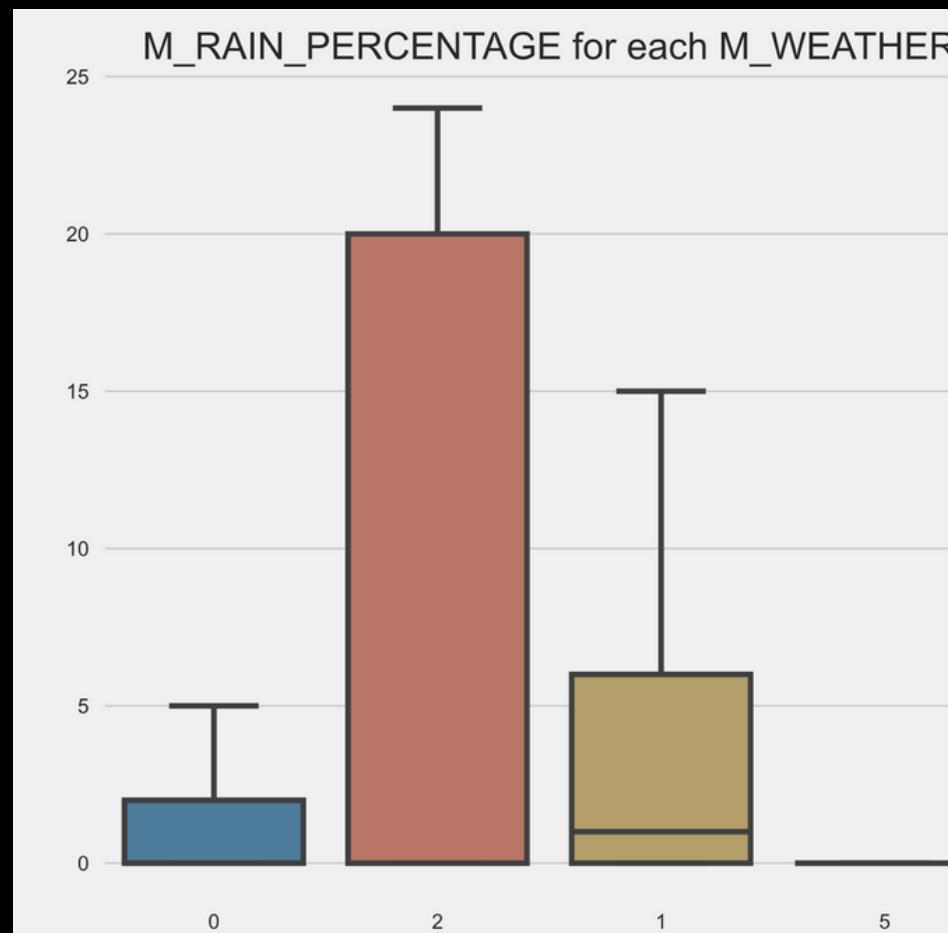
# EDA

[Go Back to Agenda Page](#)



# BiVariate Analysis

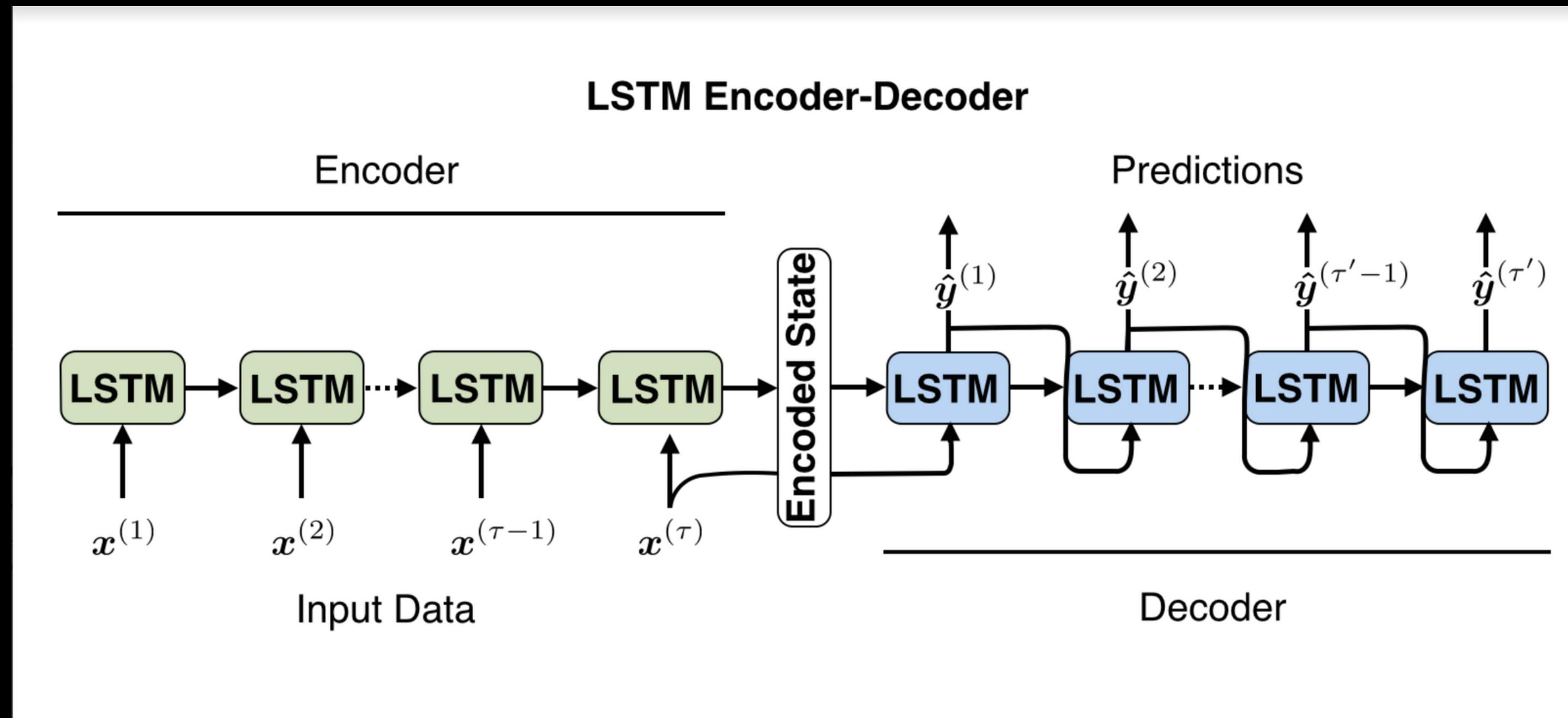




# MODEL

[Go Back to Agenda Page](#)

# MODEL WORKFLOW



Sequence-to-sequence learning (Seq2Seq) is about training models to convert sequences from one domain to sequences in another domain.

A

r

A RNN layer (or stack thereof) acts as "encoder": it processes the input sequence and returns its own internal state. Note that we discard the outputs of the encoder RNN, only recovering the state. This state will serve as the "context", or "conditioning", of the decoder in the next step.

M

h

o

i

d

t

Another RNN layer (or stack thereof) acts as "decoder": it is trained to predict the next characters of the target sequence, given previous characters of the target sequence.

e

e

Specifically, it is trained to turn the target sequences into the same sequences but offset by one timestep in the future, a training process called "teacher forcing" in this context.

l

t

Importantly, the encoder uses as initial state the state vectors from the encoder, which is how the decoder obtains information about what it is supposed to generate.

u

r

Effectively, the decoder learns to generate  $\text{targets}[t+1\dots]$  given  $\text{targets}[\dots t]$ , conditioned on the input sequence.

e

# RESULTS & CONCLUSION

[Go Back to Agenda Page](#)

- Our model was trained for the given data and MAE was found to be 0.004.
- This exhibits that our model is performing exemplary and with some modifications can be used further down the line.

### Future Work:

- Pair a CNN layer with LSTM cells and fine-tune on radar imaging weather data to be more accurate per unit area
- More cleaning of data can be done with further preprocessing, more categories can be created and resultant can be used to classify the weather type..

# FUTURE GOALS

The idea can be taken forward and an improved system can be made which can be of critical help. Hope to work on it in coming time.





THANK  
YOU!