# Things to Consider

Lookahead Bias

* Look for resources to avoid lookahead bias
* **Recommendation:** Add testing code to ensure feature generation doesn’t have this bug

Function Convention

* “predict”: sklean convention: (i) ypred is the returned object; (ii) ypred is an np.array; (iii) ypred has same shape as the data in path\_to\_test\_csv; (iv) Nth row in ypred is the prediction for the Nth row in the input matrix i.e. the order of the rows is preserved
* Deliver code as in model template

Writeup

* Feedback Questions: Keep in mind and work accordingly
* Modeling Thought Process: Keep in mind and work accordingly

# Resources

* Fyde Code
* My code present somewhere else
* Quantiacs
* Mayank Code

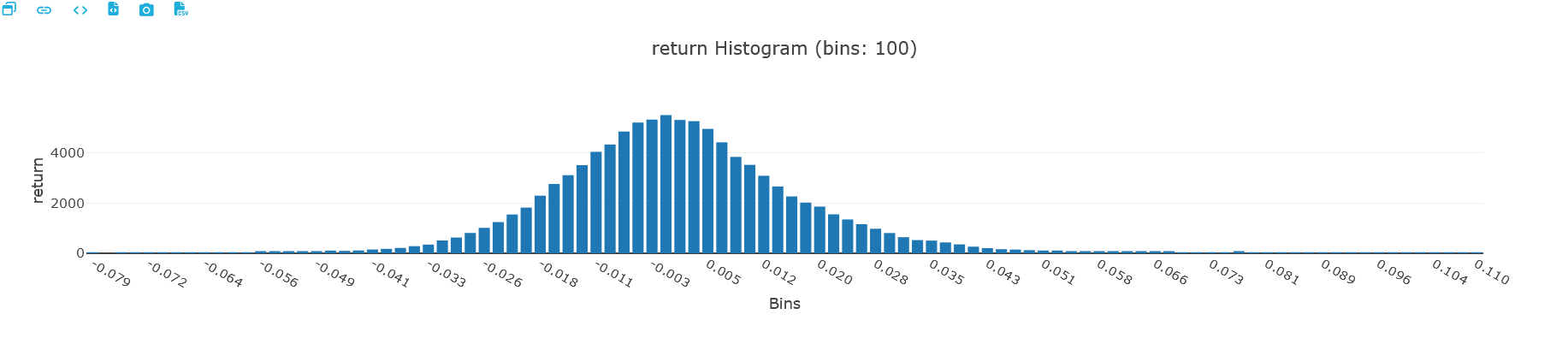
# Tasks

Understanding Data & Analysis

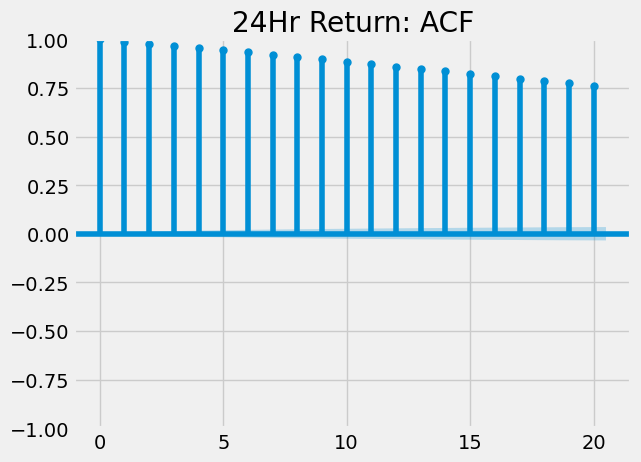
Returns Distribution

A blue sound wave with black text

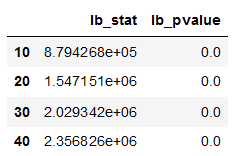
Description automatically generated



Return Properties

 A graph with blue lines

Description automatically generated



* Ljung Box Test Statistic rejects null hypothesis that return is white noise => scope of prediction
* From ACF and PACF plot, looks like AR(1) but explicit modeling below

Thus, no serial correlation and hence the traditional ARMA models won’t help in predicting the series

Base Model

Data Preparation

* Create Target Variable:
  + Exact times unavailable
  + Holiday Treatment

Modeling