

List of my best posts – 2021

- 1) Things that can get you rejected in a Data Science Interview - <https://bit.ly/3EHlb1m>
- 2) What do recruiters look for while hiring Data Science Interns? - <https://bit.ly/3eysVI5>
- 3) Why saying "We accept the Null Hypothesis" is wrong - <https://bit.ly/3qqxbiu>
- 4) Logistic Regression is Regression (meme) - <https://bit.ly/3qsHhiV>
- 5) The many names of Dependent and Independent variable - <https://bit.ly/33Uwfv6>
- 6) Most Important Assumption Checks of Linear Regression - <https://bit.ly/3eu47ky>
- 7) First principles thinking in Data Science. - <https://bit.ly/3Fx4Ey2>
- 8) Entropy and Getting Hired. - <https://bit.ly/32Na3CQ>
- 9) AutoML dangers: Abstraction and Data Science - Not a great combination!! - <https://bit.ly/33Y84ff>
- 10) Fallacy of "Modelling will take only 20% time" - <https://bit.ly/3sK5W5c>
- 11) Be wary of Automated Feature Selection – Chi Square test of Independence - <https://bit.ly/3ENhsPT>
- 12) Central Limit theorem does not kick in at $n=30$ - <https://bit.ly/3pyYVCa>
- 13) Data Science is a team effort - <https://bit.ly/3px3BbO>
- 14) Feature Augmentation - <https://bit.ly/3z8HDiK>
- 15) Lost in translation - <https://bit.ly/3pxrVKA>
- 16) Not knowing how the ML algorithms works under the hood leads to acceptance of 'default' - <https://bit.ly/3FuIEE1>
- 17) At what stage should a Data Scientist be roped in? - <https://bit.ly/3Jnmo14>

18) Descriptive statistics & good data visualizations are your two best friends -

<https://bit.ly/3FC6TQC>

19) Logistic Regression: Classification if you use Python, Regression if you use R !! -

<https://bit.ly/32JdfPu>

20) Errors and Residuals - <https://bit.ly/3JjTcbz>

21) Normal Distribution may be popular distribution but it is not the most prevalent. -

<https://bit.ly/3EySvYa>

22) Is correlation not causation? - <https://bit.ly/3Fz5dHC>

23) There never was a Normal Distribution - <https://bit.ly/3ewMa4W>

24) Linear regression does not model the raw values of Dependent Variable -

<https://bit.ly/3mGmHKN>

25) AutoML dangers: "Try all models" - <https://bit.ly/3euOV6C>

26) Clearing Misconceptions around Confidence Interval. – <https://bit.ly/3HgOOIn>

27) The ML algorithm is the nucleus of a Data Science project - <https://bit.ly/3mEmQOG>

28) Deep Learning is like Disneyland - <https://bit.ly/316TQrB>

29) Real Statisticians / Data Scientists don't test for normality. - <https://bit.ly/3pzlert>

30) All models are wrong, some are useful" ≠ "Modeling is a futile exercise -

<https://bit.ly/3z6E9wT>

31) No Full stack Data Scientist - <https://bit.ly/3sDDr9c>

32) The problem of learning "Formula First" in Data Science - <https://bit.ly/3z7kCMV>

33) Dunning-Kruger Effect in Data Science - <https://bit.ly/3FB4xS1>

34) The XY problem in Data Science - <https://bit.ly/3HmgsUa>

35) Stock markets can't be predicted - <https://bit.ly/3sBJ6wF>

36) There is no Poetic License in Data Science - <https://bit.ly/3FGrwLB>

- 37) β -hat vs Y-hat problem - <https://bit.ly/3mFqdFg>
- 38) Why the word AI - <https://bit.ly/3sCqY5O>
- 39) Assumption Selection - <https://bit.ly/3EBIOZ9>
- 40) German Tank Problem - <https://bit.ly/345BHvm>
- 41) Clearing Confusion around Skewness of Probability Distribution - <https://bit.ly/32EuNwj>
- 42) What is the harm in calling Logistic Regression a classification algorithm? - <https://bit.ly/3Jo0zP2>
- 43) Why Confusion matrix is called so? - <https://bit.ly/32IWLH7>
- 44) The curious case of statisticians recommending Cross Validation. - <https://bit.ly/32EvCVV>
- 45) 'Red Pill' moments in Data Science - <https://bit.ly/3FvVulr>
- 46) Things that surprise New Data Scientists when they first step into the corporate world. - <https://bit.ly/3sEwbtO>
- 47) Intuitive Explanations to Data Science Question (Stack exchange) - <https://bit.ly/3eMGG6p>
- 48) "You don't need enough math" – Biggest lie told to Data Scientists - <https://bit.ly/3Hns0qx>
- 49) Heuristics to judge Data Science courses - <https://bit.ly/3zevvwE>
- 50) Formula 1, Kaggle and Applied Data Science - <https://bit.ly/3mE1PnB>
- 51) Know Your Data First - Time series, Cross Sectional & Panel Data - <https://bit.ly/3JmFeFH>
- 52) Lessons for Data Scientists from Zillow fallout. - <https://bit.ly/3Ji8U6O>
- 53) Market Research - A unique but efficient path to become a Data Scientist.- <https://bit.ly/346fRYD>
- 54) How Aryma Labs Internship is a little different - <https://bit.ly/3Ji9hOK>

- 55) Putting ML algo in production is akin to putting a jelly on a conveyor belt - <https://bit.ly/33Yl1Wp>
- 56) Much like we can't observe a real black hole, we can't observe true error. - <https://bit.ly/32J20Xj>
- 57) Are Estimation and Prediction the same? - <https://bit.ly/32Bk3yN>
- 58) Importance of knowing the history of an Algorithm / Statistical Technique - <https://bit.ly/3mExpC>
- 59) The pitfalls of using SMOTE- <https://bit.ly/3HkdDTS>
- 60) Compelling reasons to not use Low code Data Science libraries. - <https://bit.ly/3sFjzT7>
- 61) Parallels Between Data Dredging and Algorithm Dredging. - <https://bit.ly/3EyBcGD>
- 62) You can google how to code but can't google which statistical technique to use - <https://bit.ly/3quDBNt>