**AI Fundamentals**

* + **Course 1** : Initial ramp up
    - Setup python env (local, COLAB) - cloud platforms for AI
    - Python core - language for DS (Playlist)
    - Essential Python Libraries (NumPy, PANDAS, SciPy) (Playlist)
    - Visualization tools (matplotlib, SEABORN) (Playlist)
    - datasets
      * Set of 300+ open source data files (CSVs, text, excel)
      * Sklearn datasets overview
      * Synthetic datasets (make\_blobs, make\_classification)

* + **Course 2 -** 
    - Statistics Essentials - Primer (Playlist)
      * Essential stats for DS : why stats and maths - overview
      * Foundations of Statistical Understanding : Data Types, Tables, and Feature Types Explored
      * Exploring Simple and Advanced Sampling Strategies: with Python Demos
      * Central Measures in Data: From Basics to Winsorizing
      * From Basics to Code: Exploring Data Dispersion Measures
      * Data Distributions: A Comprehensive Guide, with hands on python code
      * Exploring Statistical Measures: Kurtosis, Skewness, and Symmetry
      * Covariances in feature engineering (data science/ machine learning.
      * Correlations and Multi-collinearity in Feature Engineering
      * Exploring Correlation Measures in Data Science

* + Understand sense of DATA and pre-processing (Playlist)
    - Making sense of DATA for ML/DL modeling
    - Numeric Insights: Basic sanity check for Data Analysis
    - Beyond NaN: Understanding and Tackling Missing Data in Python - part 1
    - Missing values (Part 2) : Nearest Neighbor-Based Interpolation with scikit-learn
    - Outlier & Cardinality assessment : Python Code Demos and Strategies
    - A Hands-On Exploration of Data Encoding Methods
    - Practical Guide to Implementing Data Scaling Techniques in Python
    - Overcoming Data Imbalance: The Role of SMOTE in Machine Learning
    - Enhancing ML Model Generalization: Best Practices in Data Splitting

* + **Course 3**
    - Machine Learning - Primer (Playlist)
      * Machine Learning: Concepts, Models, and Platform Insights
      * ML - intuitive understanding
        + Supervised classification
        + Supervised regression
        + Unsupervised learning
      * Distance and similarity measures - Primer
        + Euclidean Distance        - Mostly used for quantitative data
        + Taxicab Geometry        - Used when the data types are heterogenous
        + Minkowski distance        - Intended for real-valued vector spaces
        + Jaccard index         - Often used in applications when dealing with binarized data
        + Hamming distance        - Typically used with data transmitted over computer networks. And also used with categorical variables.
        + Levenshtein Distance (Edit Distance)
        + Canberra Distance
        + Chebyshev Distance (Infinity Norm)

* + KNN- concept
  + KNN - Choosing the Right 'k'
  + KNN - Break the ties (prediction)
  + KNN - Importance of scaling the data
  + KNN - Handling Categorical Data
  + KNN - Model evaluation methods
  + KNN - Tuning for performance
  + KNN - Shortcomings in KNN
  + KNN - Saving/loading the model
  + KNN - KNN as regressor

* + LinReg - basic intuition (code)
  + LinReg - statistical way (code)
  + LinReg - sklearn implementation on advertising dataset
  + LinReg - Model evaluations (learning curve and cross validations)
  + LinReg - Test of assumptions (adv dataset)
  + LinReg - MSE plot
  + LinReg - Save/load model
  + LinReg - effect of OHE
  + LinReg - effect of multi-collinearity
  + LinReg - with non linear data

* + Unsupervised models - K-MEANS - concepts
  + K-MEANS - implementing (good data, messy data)
  + K-MEANS - limitations
  + K-MEANS - variants

* + Statistical Inference - From Hypotheses to Bayes (Playlist)

1. Hypothesis Testing

3. Bayesian Statistics