Data Science:

General:

Series:

**Dictionaries**: Structure that takes in a key and returns a value.. [{}, {}]

**Tuples:** A type of list that cannot be modified.. Can’t add or remove to elements ()

**List: {}**

**Sets:**  A type of dictionary with only keys and no values which stores unique objects only.

**Hashable:** This includes ints, floats, strings, tuples.. that are immutable

**Unhashable:** Includes dictionaries, and libraries used for larger set.

Chapter 3:

**Dataframes:**

A central object in panda with table containing rows and coluns

**Series:** An array of objects with all same type with an index

Chapter 5:

5.6:

Means, Standard Deviations, Medians and Quantiles.

* Assume that data distribution is bell shaped.

Standard Deviation: A quantity expressing by how much the members of a group differ from the mean value for the group. This measures how spread out numbers are. If a spread has a low SD, it shows it is not well spread.. The higher its spread, the higher the SD.

Quantiles:

0 quantile = 0 percentile

.25 quantile = 25 percentile

0.5 quantile = 50 percentile

0.75 quantile = 75 percentile

1 quantile = 100 percentile

Chapter 5:

5.9 Scatterplots and Logarithmic Axes:

The application of Logarithmic axes helps users to focus on the major areas in the scattered plot graph.

5.11: Heatmaps:

This shows a better representation over scatterplots, majorly good when your points are exactly on top of each other. This tool is very good for density representation.

6.0: Machine Learning Overview:

Supervised vs Unsupervised Learning:

Supervised:

Input and output variables will be given

Learning goal is to determine the function so well that when new input data set given, can predict the output

Machine learning problems, data mining and neural network

Classification, Regression, Linear Regression, Support Vector Machine

Supervised Learning is ofter used for export systems in image recognition, speech recognition, forecasting, financial

analysis and training neural networks and decision trees

Unsupervised:

Only input data will be given

Learning goal is to model the hidden patterns or underlying structure in the given input data in order to learn about the data

Machine learning problems, data mining and neural network

Clustering, Association, k-means, Association

Unsupervised learning algorithms are used to pre-process the data, during exploratory analysis or to pre-train supervised

learning algorithms

**Chapter 7**:

Interlude: Feature Extraction Ideas:

7.1: Standard Features:

Is\_null: Used for null entry,.. incase data is missing

Dummy Variable: