**MACHINE LEARNING**

**METHODOLOGY IN MACHINE LEARNING:**

1. Identify Problem
2. Gathering Data
3. EDA (Explore Data Analysis)
4. Built Model and Train & Develop Model
5. Move into Production

**EDA ANALYSIS:**

EDA i.e. Exploratory Data Analysis is the process of studying data by leveraging various statistical and visualization techniques

1. Univariate Analysis
2. Bivariate Analysis
3. Multivirate Analysis

Univariate Analysis : Impact of one variable in dataset….Study of single variable at one point of time

Ex: marks of students

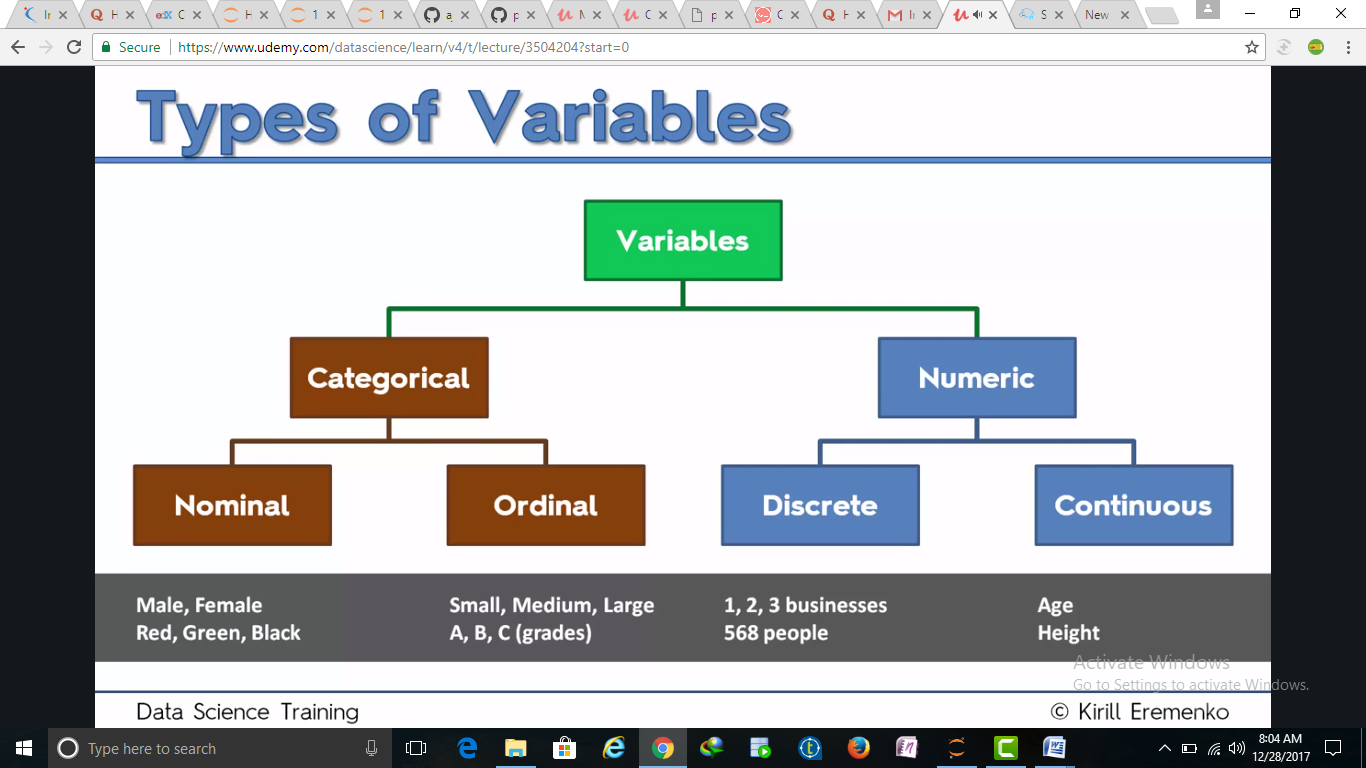
Marks is a variable for student dataset

* There are two types of variables :

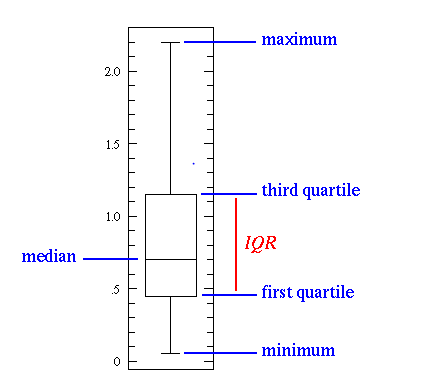
1. Categorical ----Strings
2. Numerical ------Numbers

**Plots on Categorical Variables : Bar chart & Pie Chart**

**Plots on Numerical Variables: Box Plot & Histogram**

****

**BOX PLOT:**

****

Histogram:

Variance= represents data dispersion

Standard deviation : sqrt(Variance)

Co-efficient deviation : Standard Deviation/mean

Skewness : represents symmetry of graph ,if it is Inclined right means +ve symmetry

Kurtosis: Whether data is dispersed peaked or flatted

Random =[2,4,6,8,10]

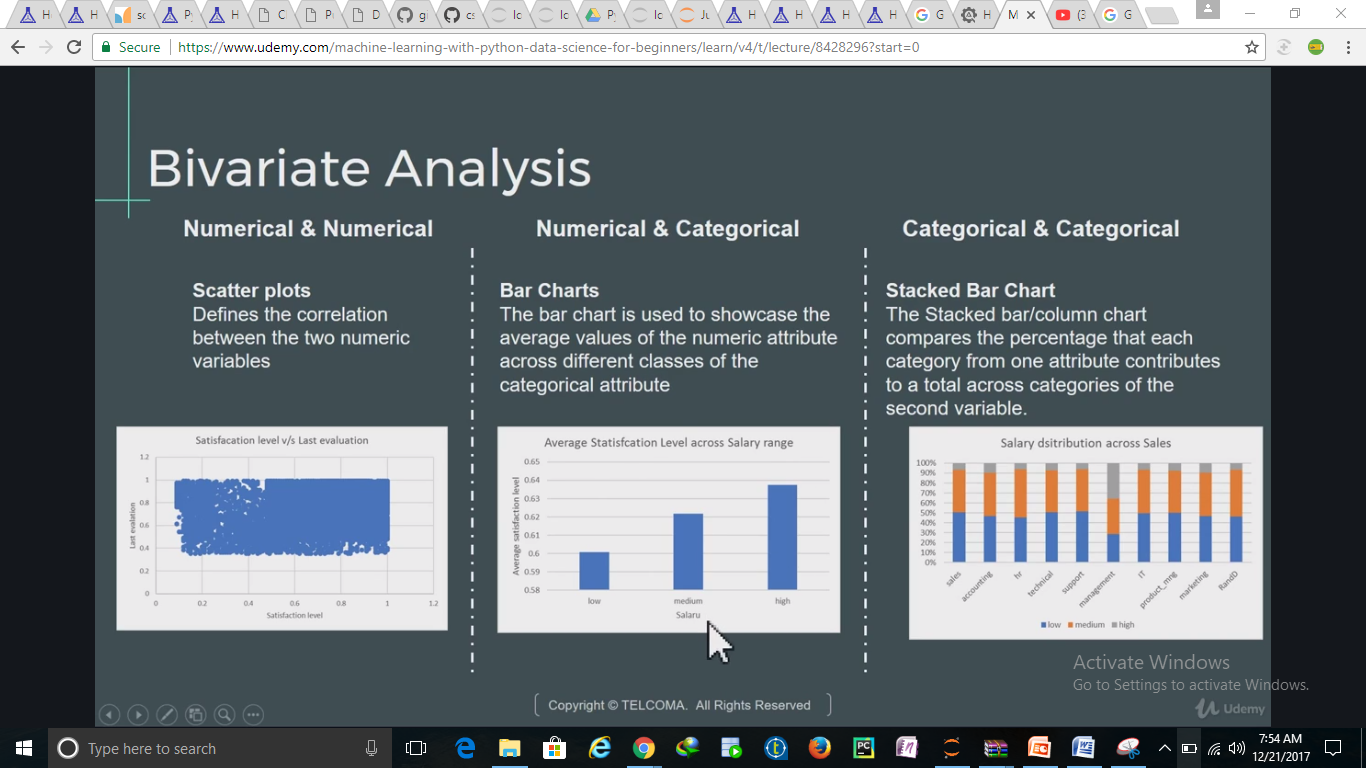
Variance = (sum (random)^2/N-1

Mean = sum/num

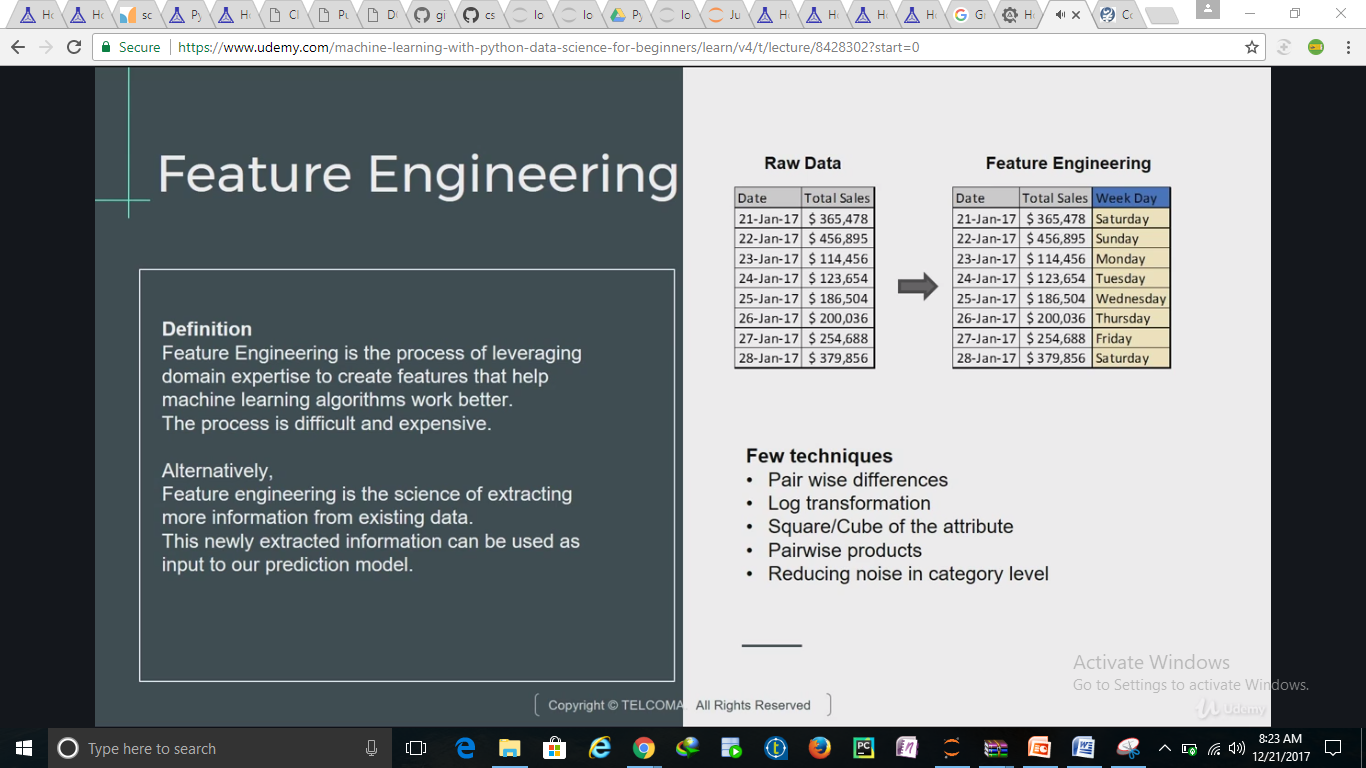
S.D =sqrt(Variance)

Coefficient = S.D /mean

Bivariate Analysis:



Feature Engineering:



Extracting weekday from dates makes sense

**Basics of Statistics:**

**Mean**: sum of quantities/ number of quantities (AVG)

**Median**: central point or mid value

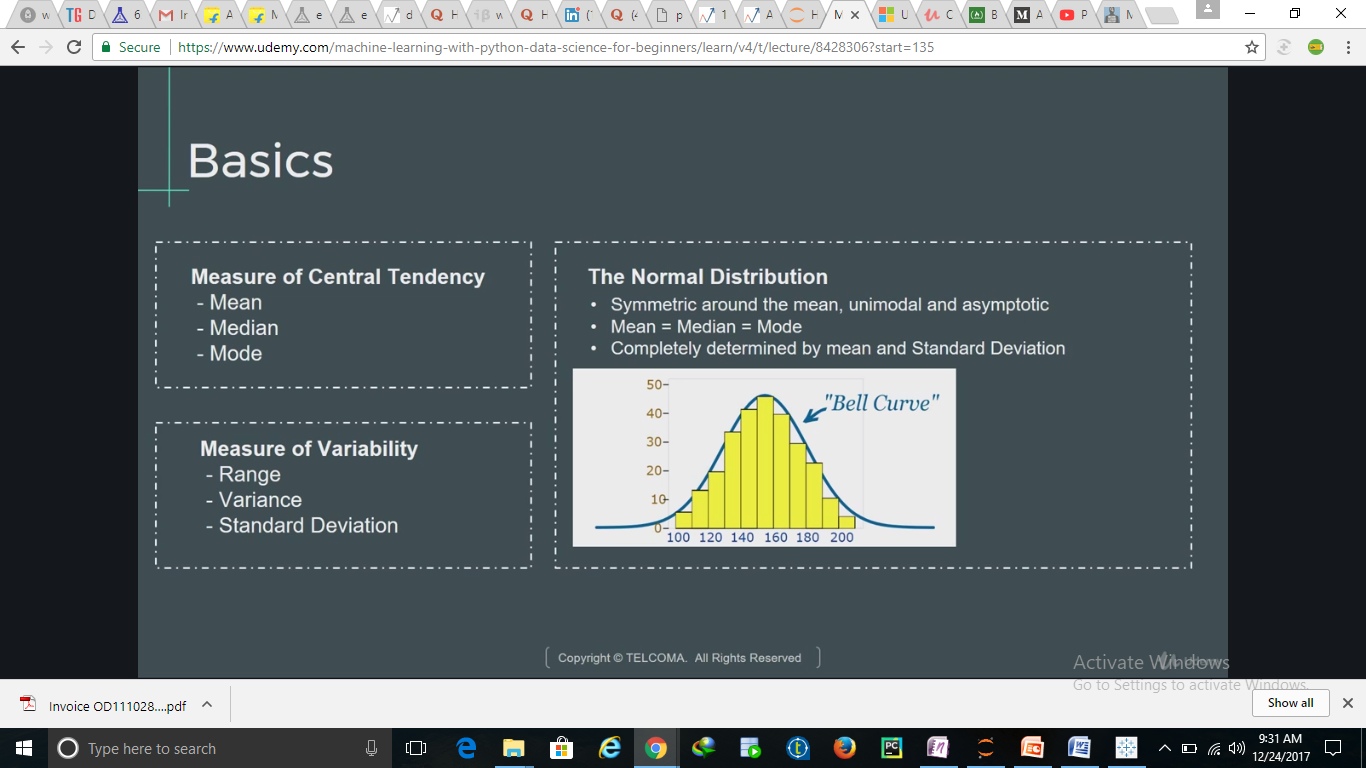
**Mode** : most occurrence value

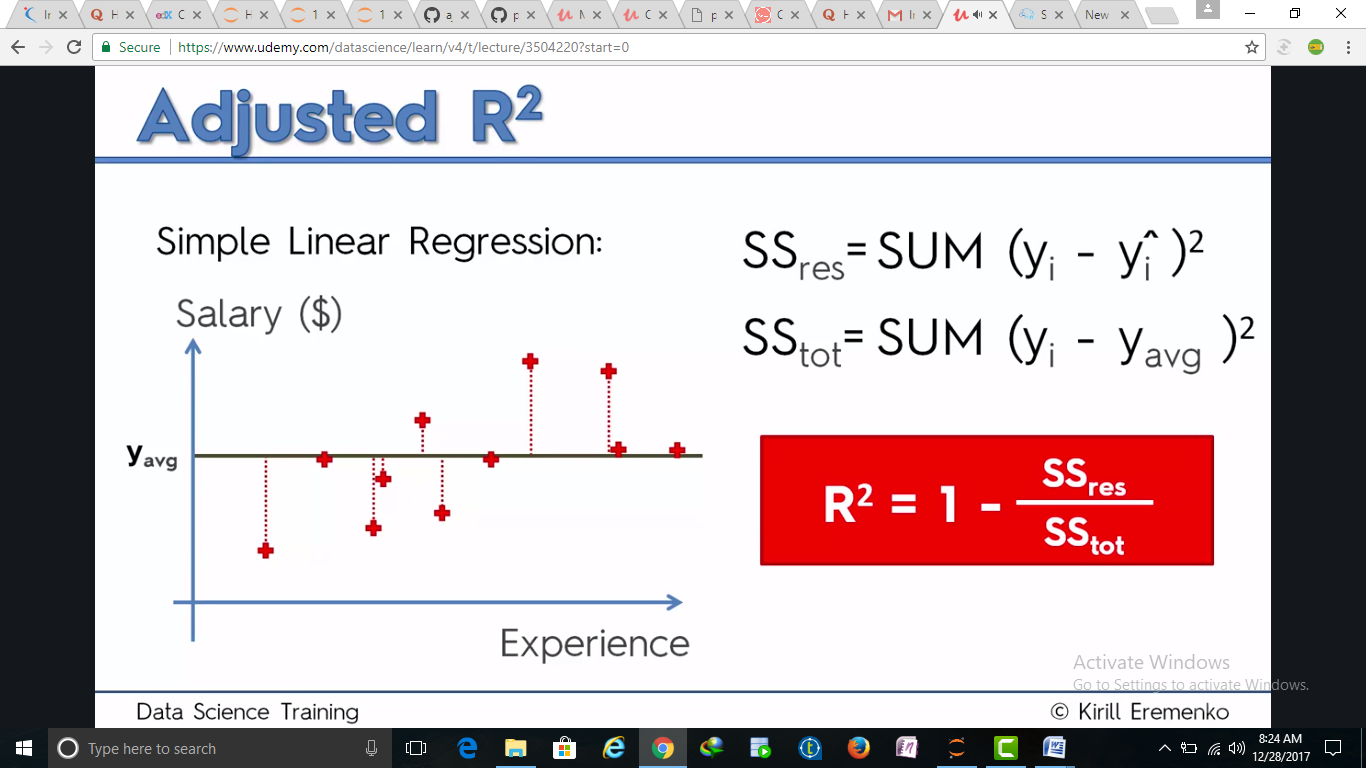
**Range:** Starting Point and Ending point

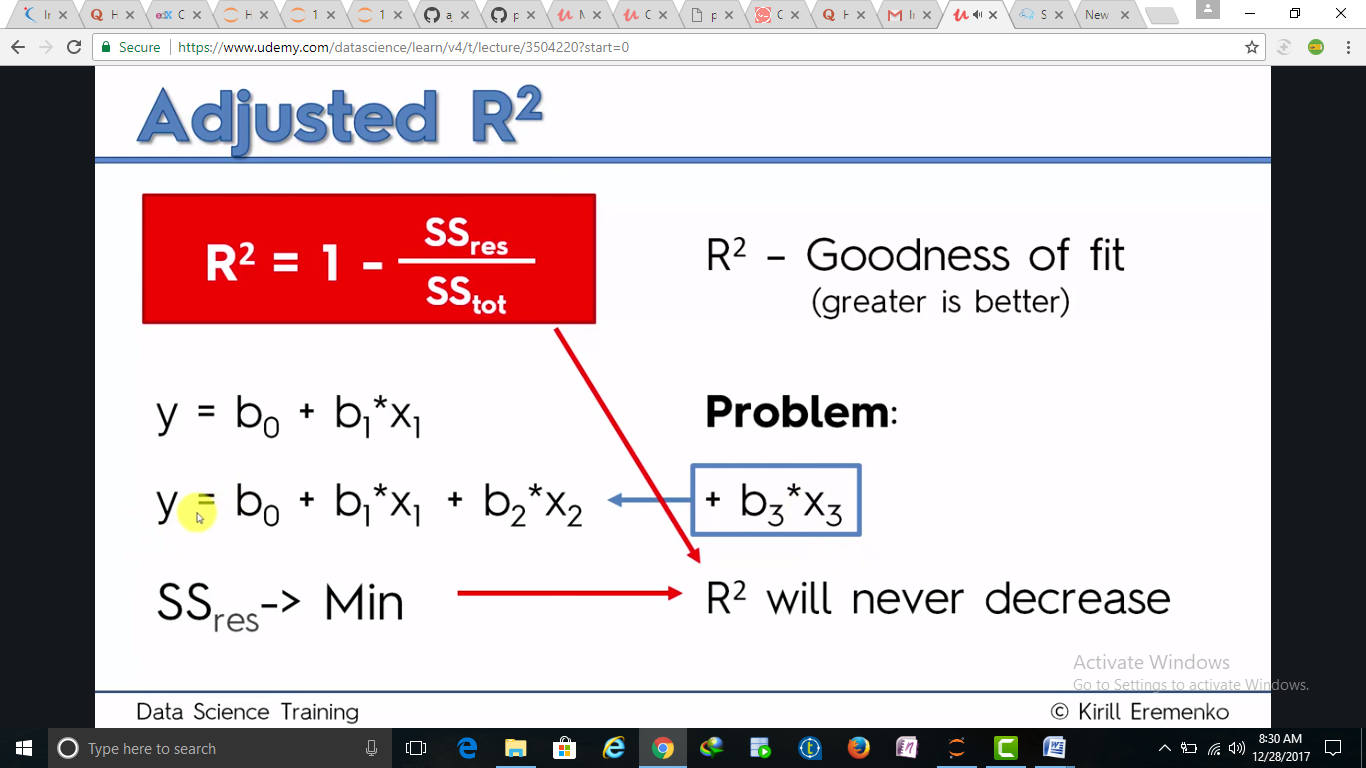
**Variance:** Distribution of points

**Standard deviation:** Sqrt(Variance)

**Normal Distribution:** Mean=Median =Mode

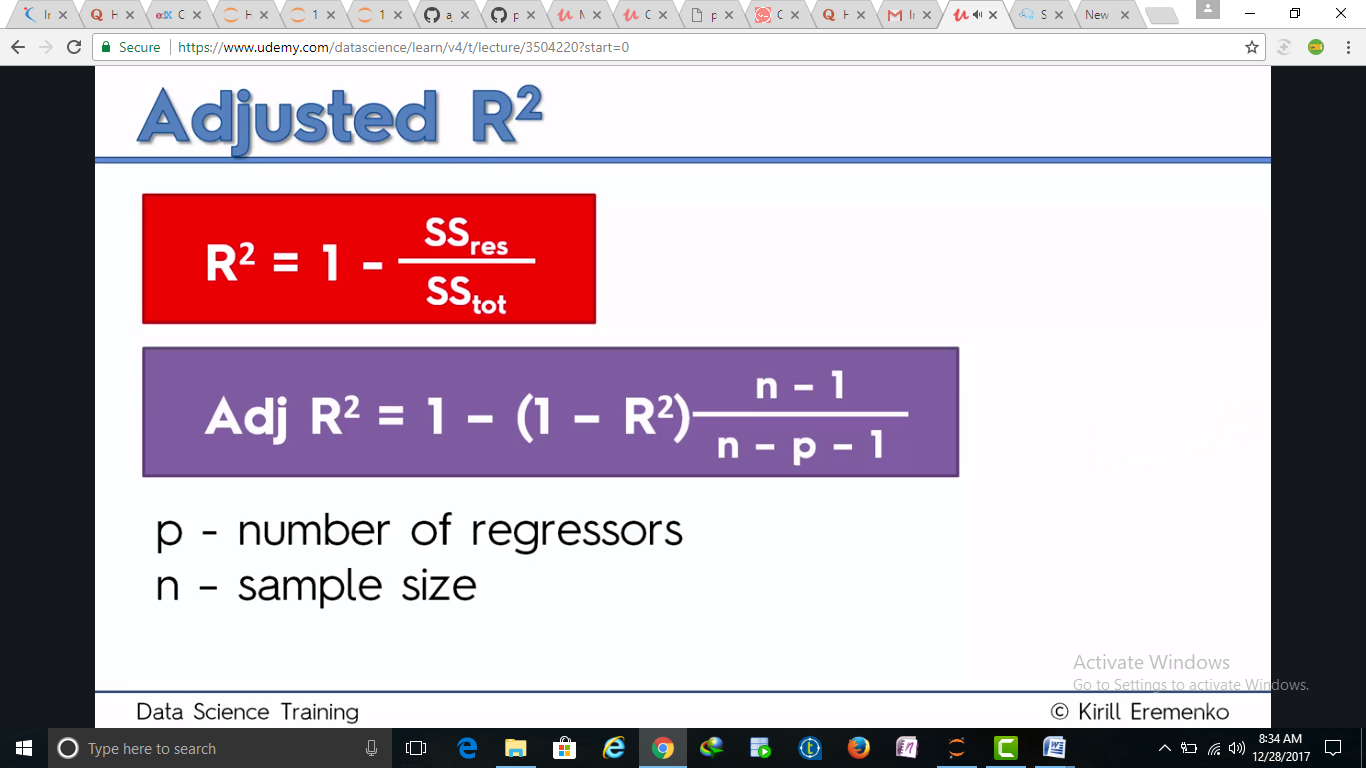
****

****

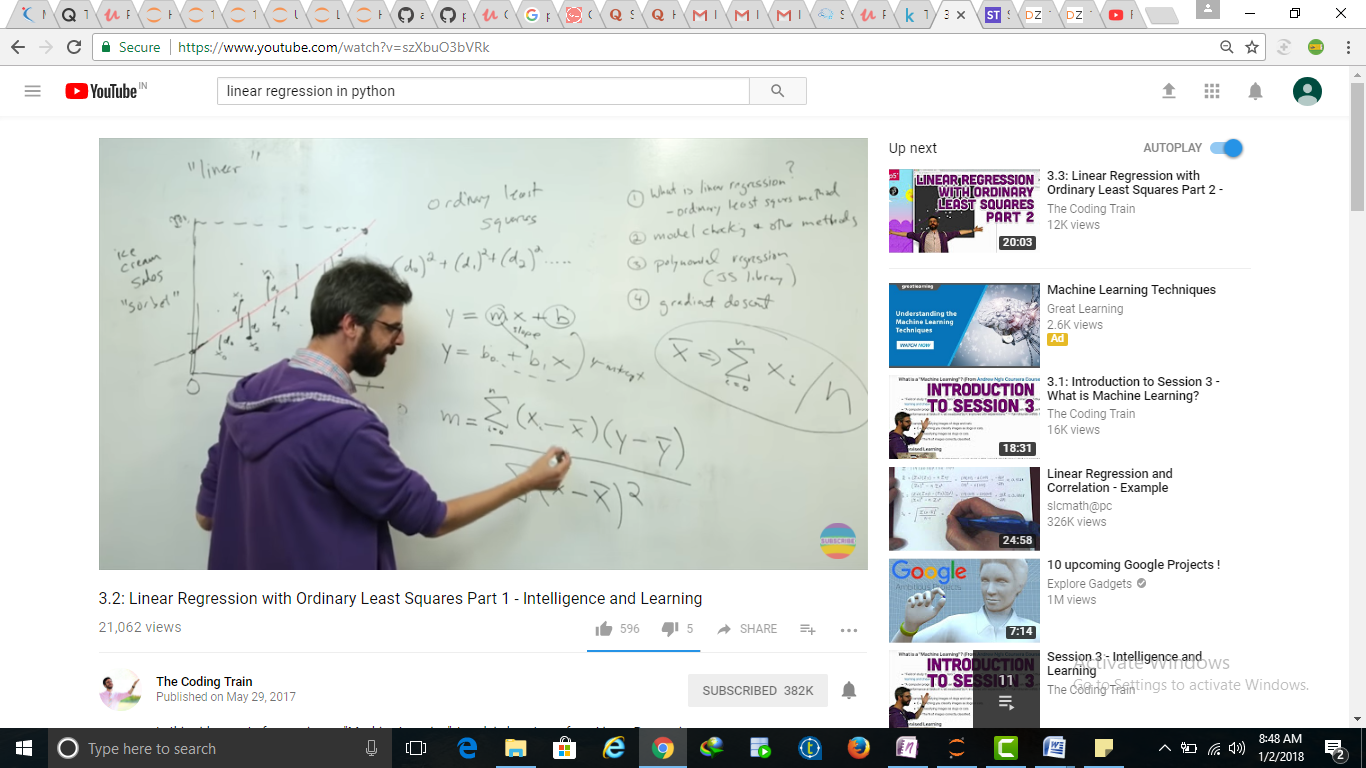
****

R2 never decrease by adding variables .bcoz ss res is always minimum

R2 should be close to 1



Adj R2 gives whether you are adding good variables or not



X bar =Avg of x to cal Slope m

