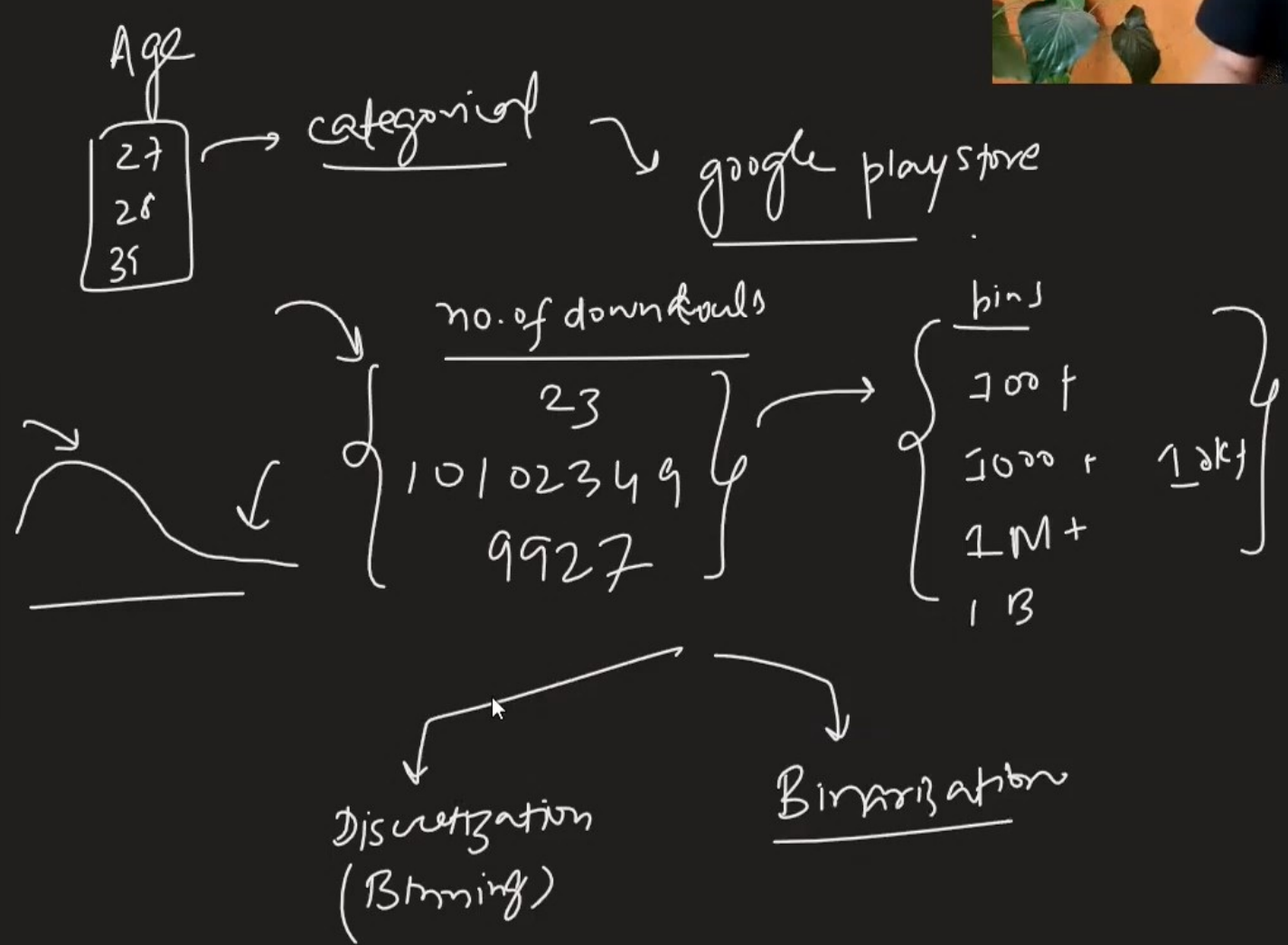


100 Days of ML	1. Encoding Numerical Features
Day 14 - Framing t...	2. Discretization
Day 15 - Working...	3. Types of Discretization
Day 16 - Working...	4. Equal Width/Uniform Binning
Day 17 - API to Pa...	5. Equal Frequency/Quantile Binni...
Day 18 - Web Scra...	6. KMeans Binning
Day 19 - Understa...	7. Encoding the discretized variable
Day 20 - Univariate...	8. Example
Day 21 - Bivariate...	9. Custom/Domain Based Binning
Day 22 - Pandas Pr...	10. Binarization
Day 23 - Feature E...	11. Example
Day 24 - Standardi...	
Day 25 - Normaliza...	
Day 26 - Ordinal E...	
Day 27 - One Hot E...	
Day 28 - ColumnTr...	
Day 29 - Pipelines	
Day 30 - Function T...	
Day 31 - Power Tra...	
Day 32 - Discrtizati...	
+ Add section	+ Add page

# 1. Encoding Numerical Features

Monday, April 19, 2021 4:18 PM





## 2. Discretization

Discretization is the process of transforming continuous variables into discrete variables by creating a set of contiguous intervals that span the range of the variable's values. Discretization is also called binning, where bin is an alternative name for interval.

Why use Discretization:

1. To handle Outliers
2. To improve the value spread



Age

23 42, 57 81 . . . 100

↓

0-10, 10-20, 20-30 . . .

5          6          10

100 Days of ML

Day 14 - Framing t...

Day 15 - Working...

Day 16 - Working...

Day 17 - API to Pa...

Day 18 - Web Scra...

Day 19 - Understa...

Day 20 - Univariate...

Day 21 - Bivariate...

Day 22 - Pandas Pr...

Day 23 - Feature E...

Day 24 - Standardi...

Day 25 - Normaliza...

Day 26 - Ordinal E...

Day 27 - One Hot E...

Day 28 - ColumnTr...

Day 29 - Pipelines

Day 30 - Function T...

Day 31 - Power Tra...

Day 32 - Discrtizati...

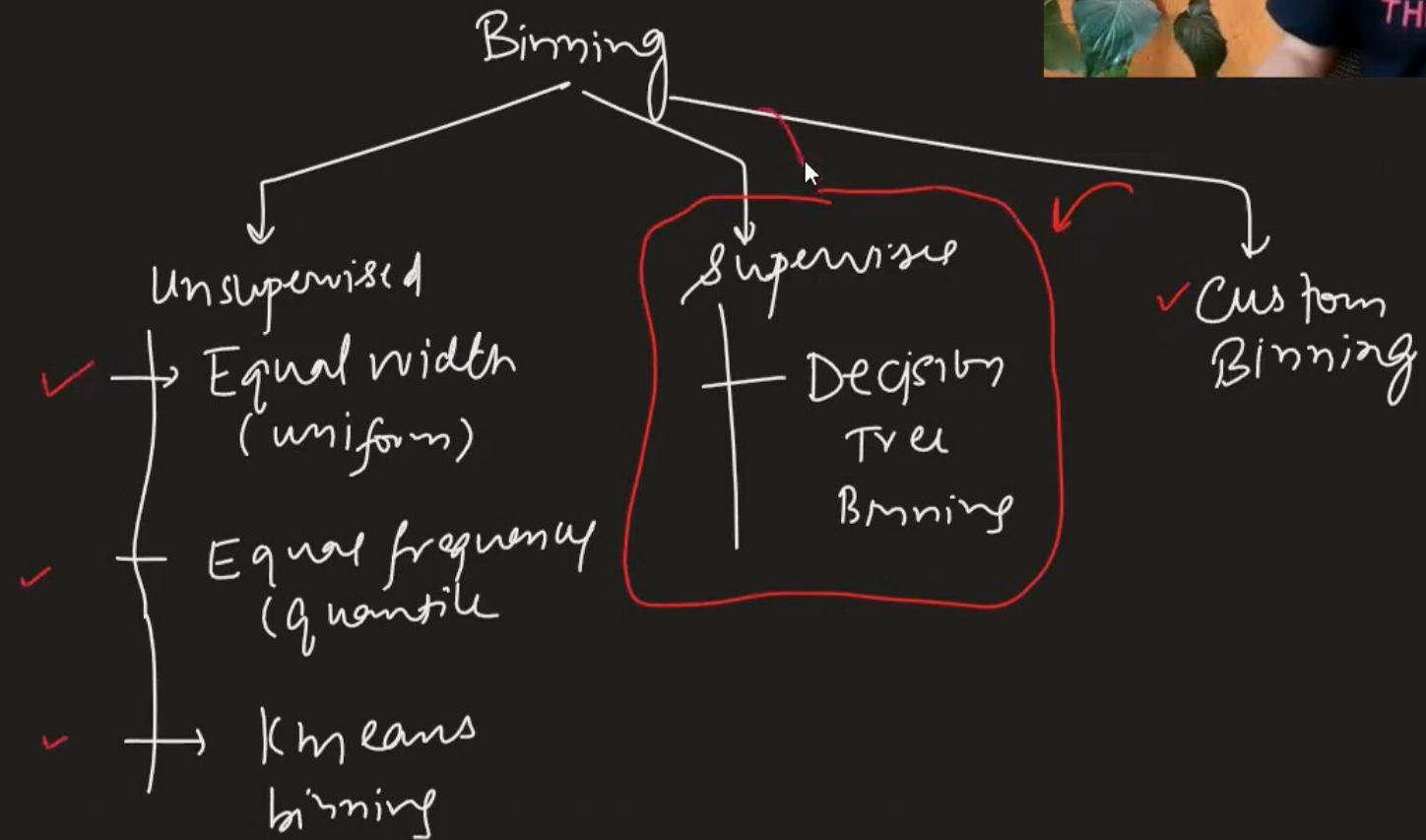
+ Add section

+ Add page



### 3. Types of Discretization

Monday, April 19, 2021 3:56 PM



100 Days of ML

Day 14 - Framing t...  
Day 15 - Working...  
Day 16 - Working...  
Day 17 - API to Pa...  
Day 18 - Web Scra...  
Day 19 - Understa...  
Day 20 - Univariate...  
Day 21 - Bivariate...  
Day 22 - Pandas Pr...  
Day 23 - Feature E...  
Day 24 - Standardi...  
Day 25 - Normaliza...  
Day 26 - Ordinal E...  
Day 27 - One Hot E...  
Day 28 - ColumnTr...  
Day 29 - Pipelines  
Day 30 - Function T...  
Day 31 - Power Tra...  
Day 32 - Discrtizati...

1. Encoding Numerical Features
2. Discretization
3. Types of Discretization
4. Equal Width/Uniform Binning
5. Equal Frequency/Quantile Binni...
6. KMeans Binning
7. Encoding the discretized variable
8. Example
9. Custom/Domain Based Binning
10. Binarization
11. Example

+ Add section

+ Add page



← → OneNote for Windows 10

Home Insert Draw View Help

↶ ↷ AI + ⇄

📁 100 Days of ML

- Day 14 - Framing t...
- Day 15 - Working...
- Day 16 - Working...
- Day 17 - API to Pa...
- Day 18 - Web Scra...
- Day 19 - Understa...
- Day 20 - Univariate...
- Day 21 - Bivariate...
- Day 22 - Pandas Pr...
- Day 23 - Feature E...
- Day 24 - Standardi...
- Day 25 - Normaliza...
- Day 26 - Ordinal E...
- Day 27 - One Hot E...
- Day 28 - ColumnTr...
- Day 29 - Pipelines
- Day 30 - Function T...
- Day 31 - Power Tra...
- Day 32 - Discrtizati...

1. Encoding Numerical Features

2. Discretization

3. Types of Discretization

4. Equal Width/Uniform Binning

5. Equal Frequency/Quantile Binni...

6. KMeans Binning

7. Encoding the discretized variable

8. Example

9. Custom/Domain Based Binning

10. Binarization

11. Example

+ Add section + Add page

### 4. Equal Width/Uniform Binning

Monday, April 19, 2021 3:56 PM

Age

1) Outliers  
2) No change spread

(27), 32, 84, 56, ... max 100

Bins = 10

$$\frac{\text{max} - \text{min}}{\text{bins}} = \frac{100 - 0}{10} = 10$$

10 bins

(0-10), (10-20), (20-30), ... (90-100)

5 16 17 1 1 1 5

equal binning



	age	age_trf	age_labels
314	<u>43.0</u>	<u>5.0</u>	( <u>40.21</u> , <u>48.168</u> ]
523	<u>44.0</u>	<u>5.0</u>	( <u>40.21</u> , <u>48.168</u> ]
352	<u>15.0</u>	<u>1.0</u>	( <u>8.378</u> , <u>16.336</u> ]
534	<u>30.0</u>	<u>3.0</u>	( <u>24.294</u> , <u>32.252</u> ]
211	35.0	4.0	(32.252, 40.21]
530	2.0	0.0	(0.42, 8.378]
786	18.0	2.0	(16.336, 24.294]
827	1.0	0.0	(0.42, 8.378]
372	19.0	2.0	(16.336, 24.294]
518	36.0	4.0	(32.252, 40.21]



← → OneNote for Windows 10

Home Insert Draw View Help

100 Days of ML

- Day 14 - Framing t...
- Day 15 - Working...
- Day 16 - Working...
- Day 17 - API to Pa...
- Day 18 - Web Scra...
- Day 19 - Understa...
- Day 20 - Univariate...
- Day 21 - Bivariate...
- Day 22 - Pandas Pr...
- Day 23 - Feature E...
- Day 24 - Standardi...
- Day 25 - Normaliza...
- Day 26 - Ordinal E...
- Day 27 - One Hot E...
- Day 28 - ColumnTr...
- Day 29 - Pipelines
- Day 30 - Function T...
- Day 31 - Power Tra...
- Day 32 - Discrtizati...

1. Encoding Numerical Features

2. Discretization

3. Types of Discretization

4. Equal Width/Uniform Binning

5. Equal Frequency/Quantile Binning

6. KMeans Binning

7. Encoding the discretized variable

8. Example

9. Custom/Domain Based Binning

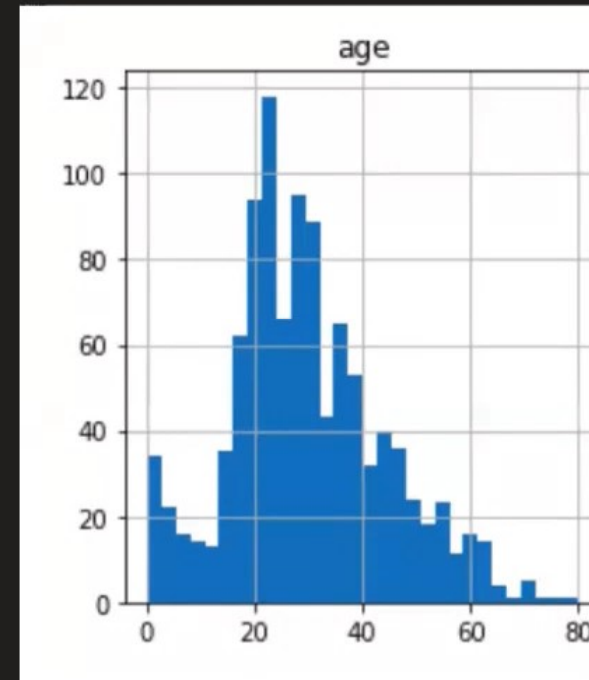
10. Binarization

11. Example

+ Add section + Add page

## 5. Equal Frequency/Quantile Binning

Monday, April 19, 2021 3:56 PM



Intervals = 10

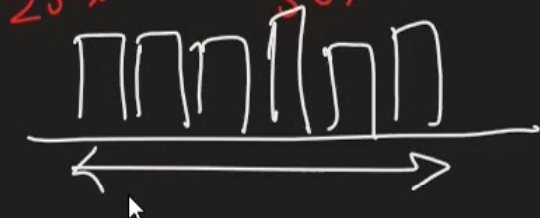
Each interval contains 10% of total observations

Intervals: 0-16; 16-20; 20-22; 22-25; ... 50-74

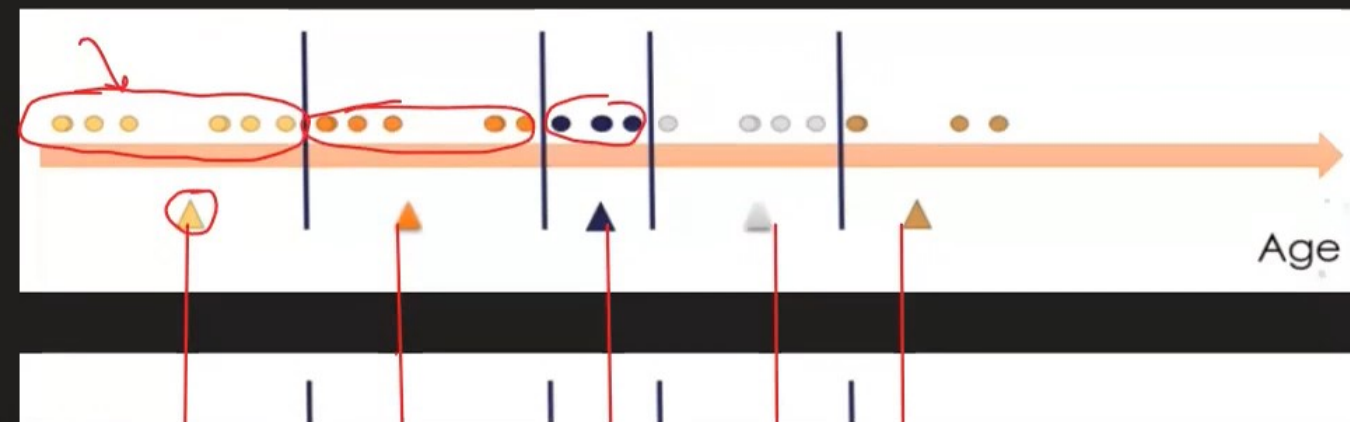
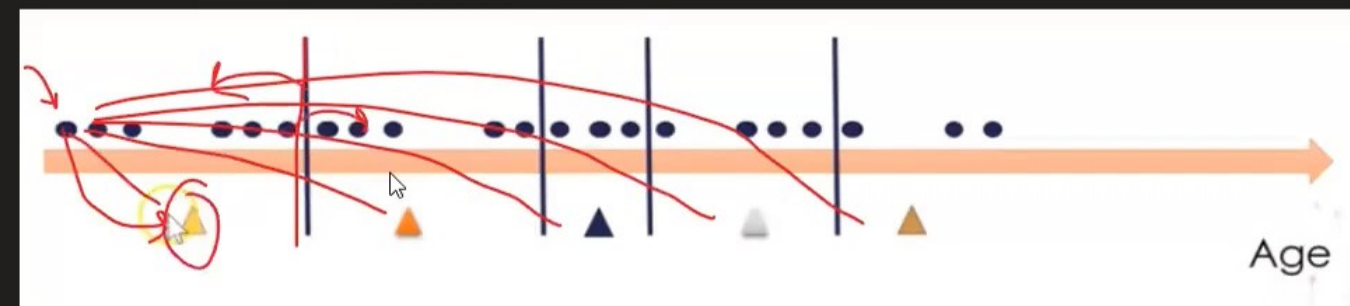
1) Outliers  
2) Value spread spread

0 - (?)  
10<sup>th</sup> person

0-16 16-20 20-22 -  
10% 20% 30%







K Means

↳ Clustering

2D

40D

1. Encoding Numerical Features
2. Discretization
3. Types of Discretization
4. Equal Width/Uniform Binning
5. Equal Frequency/Quantile Binning
6. KMeans Binning
7. Encoding the discretized variable
8. Example
9. Custom/Domain Based Binning
10. Binarization
11. Example

OneNote for Windows 10

Home Insert Draw View Help

100 Days of ML

- Day 14 - Framing t...
- Day 15 - Working...
- Day 16 - Working...
- Day 17 - API to Pa...
- Day 18 - Web Scra...
- Day 19 - Understa...
- Day 20 - Univariate...
- Day 21 - Bivariate...
- Day 22 - Pandas Pr...
- Day 23 - Feature E...
- Day 24 - Standardi...
- Day 25 - Normaliza...
- Day 26 - Ordinal E...
- Day 27 - One Hot E...
- Day 28 - ColumnTr...
- Day 29 - Pipelines
- Day 30 - Function T...
- Day 31 - Power Tra...
- Day 32 - Discrtizati...

1. Encoding Numerical Features

2. Discretization

3. Types of Discretization

4. Equal Width/Uniform Binning

5. Equal Frequency/Quantile Binni...

6. KMeans Binning

7. Encoding the discretized variable

8. Example

9. Custom/Domain Based Binning

10. Binarization

11. Example

+ Add section + Add page

## 7. Encoding the discretized variable

Monday, April 19, 2021 3:58 PM

SKlearn

↳ KBinsDiscretizer()

bins = ?

strategy

+ uniform  
+ quantile  
+ kmeans

encoding

+ ordinal  
+ onehotencoding



← → OneNote for Windows 10

Home Insert Draw View Help

↶ ↷ AI ↻ ⇅

📏 🖌️ 🖌️ 🖌️ 🖌️ 🖌️ 🖌️ 🖌️ 🖌️ + 📐 Shapes 🔗 Ink to Shape ↺ Ink to Text

100 Days of ML ▾

🔍 🕒

- Day 14 - Framing t...
- Day 15 - Working...
- Day 16 - Working...
- Day 17 - API to Pa...
- Day 18 - Web Scra...
- Day 19 - Understa...
- Day 20 - Univariate...
- Day 21 - Bivariate...
- Day 22 - Pandas Pr...
- Day 23 - Feature E...
- Day 24 - Standardi...
- Day 25 - Normaliza...
- Day 26 - Ordinal E...
- Day 27 - One Hot E...
- Day 28 - ColumnTr...
- Day 29 - Pipelines
- Day 30 - Function T...
- Day 31 - Power Tra...
- Day 32 - Discrtizati...

1. Encoding Numerical Features
2. Discretization
3. Types of Discretization
4. Equal Width/Uniform Binning
5. Equal Frequency/Quantile Binni...
6. KMeans Binning
7. Encoding the discretized variable
8. Example
9. Custom/Domain Based Binning
10. Binarization
11. Example

+ Add section + Add page

## 9. Custom/Domain Based Binning

Monday, April 19, 2021 3:58 PM

$[0 - 18] \rightarrow \text{Kids}$   
 $[18 - 60] \rightarrow$   
 $[60 - 80] \rightarrow$

Sklearn  
↳ Pandas







## 10. Binarization

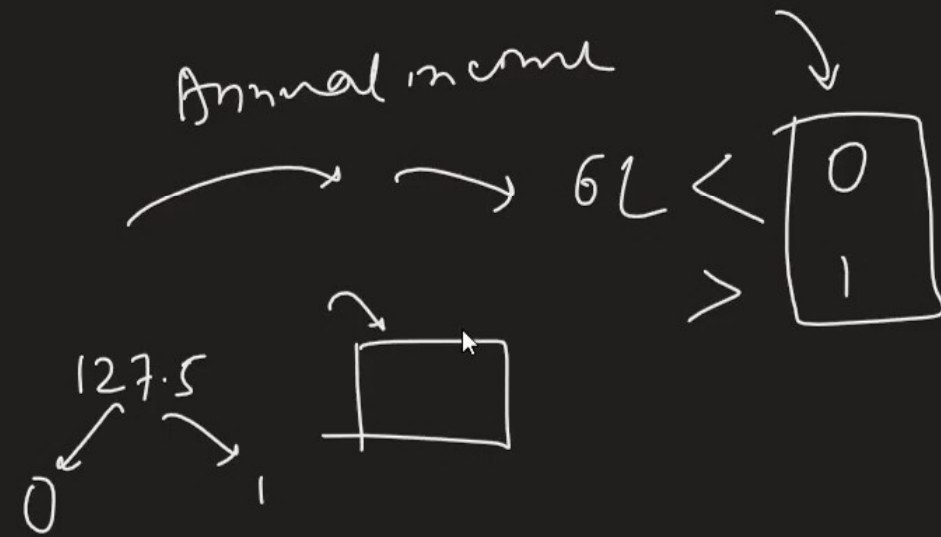
Monday, April 19, 2021 4:17 PM



image 0, 1



0-255  
color



100 Days of ML

Day 14 - Framing t...

Day 15 - Working...

Day 16 - Working...

Day 17 - API to Pa...

Day 18 - Web Scra...

Day 19 - Understa...

Day 20 - Univariate...

Day 21 - Bivariate...

Day 22 - Pandas Pr...

Day 23 - Feature E...

Day 24 - Standardi...

Day 25 - Normaliza...

Day 26 - Ordinal E...

Day 27 - One Hot E...

Day 28 - ColumnTr...

Day 29 - Pipelines

Day 30 - Function T...

Day 31 - Power Tra...

Day 32 - Discrtizati...

+ Add section

+ Add page