OneNote for Windows 10

Home Insert Draw View Help

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...

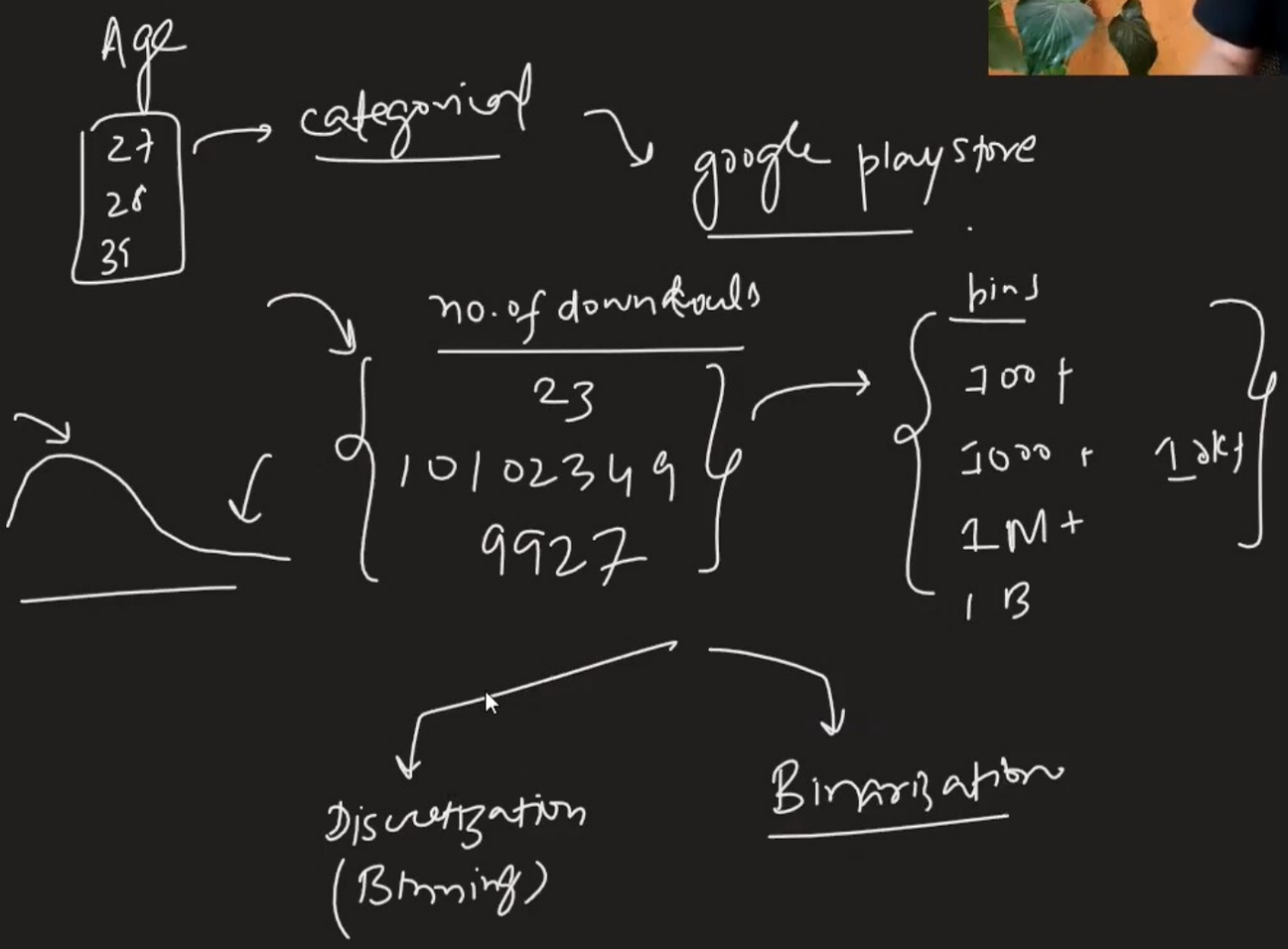
Add section

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 page

# 1. Encoding Numerical Features

Monday, April 19, 2021 4:18 PM




OneNote for Windows 10

Home Insert Draw View Help

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...

Add section

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

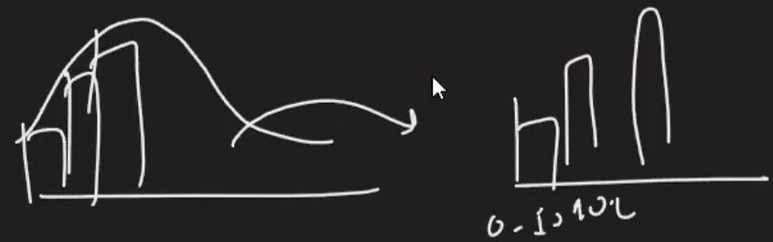
Add page

# 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

OneNote for Windows 10

Home Insert Draw View Help

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...

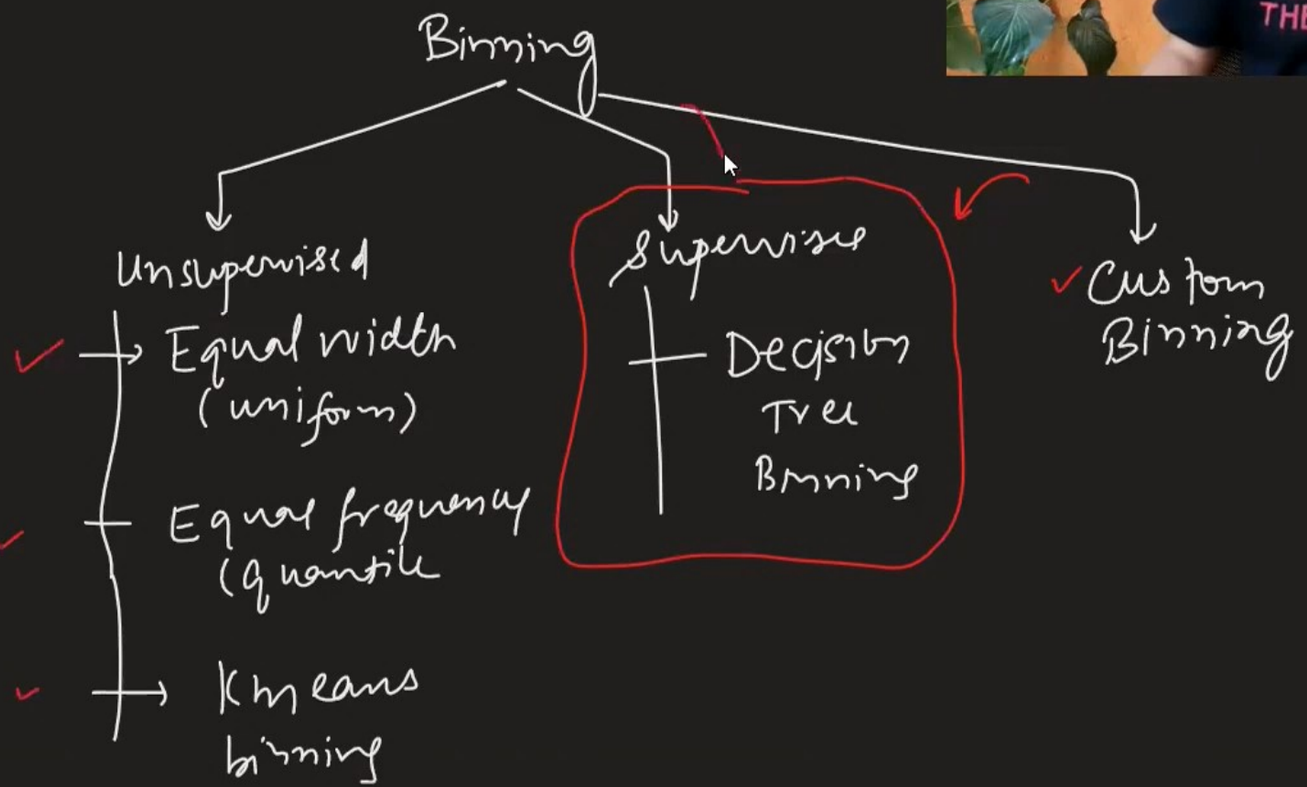
Add section

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 page

# 3. Types of Discretization

Monday, April 19, 2021 3:56 PM




OneNote for Windows 10

Home Insert Draw View Help

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...

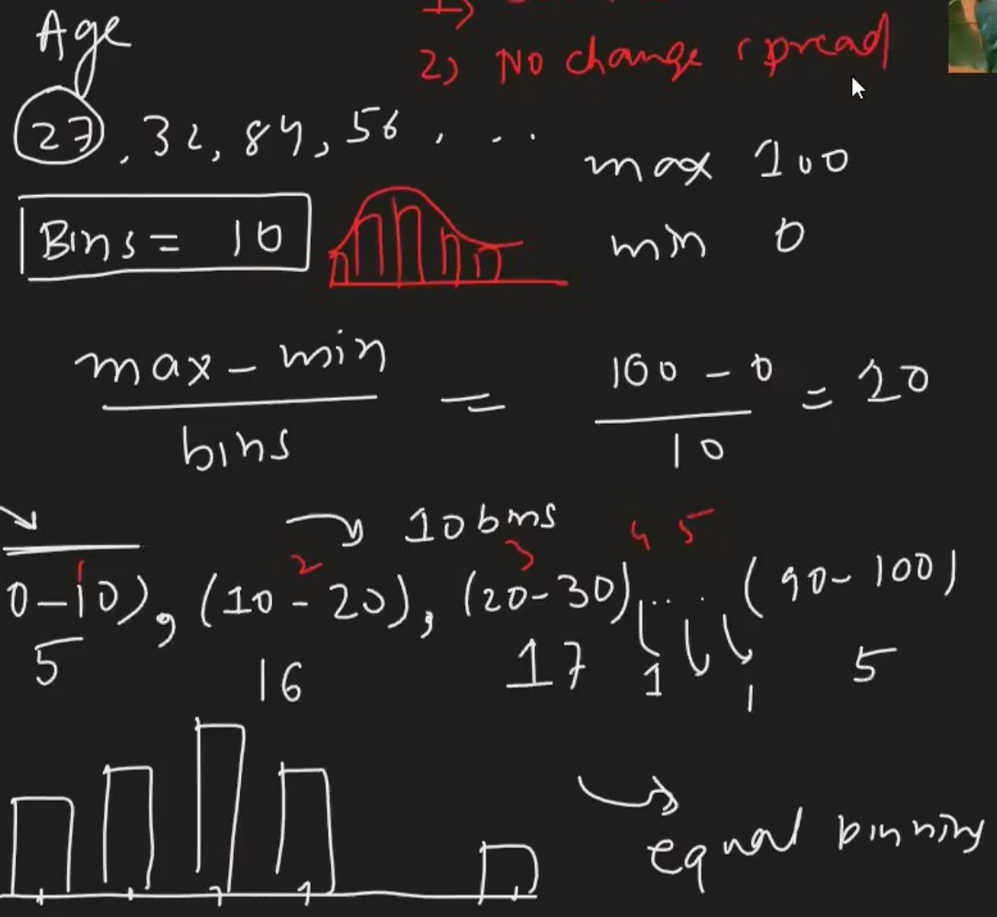
Add section

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 page

# 4. Equal Width/Uniform Binning

Monday, April 19, 2021 3:56 PM




| | age | age_trf | age_labels |
|---|---|---|---|
| 314 | 43.0 | 5.0 | (40.21, 48.168] |
| 523 | 44.0 | 5.0 | (40.21, 48.168] |
| 352 | 15.0 | 1.0 | (8.378, 16.336] |
| 534 | 30.0 | 3.0 | (24.294, 32.252] |
| 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

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...

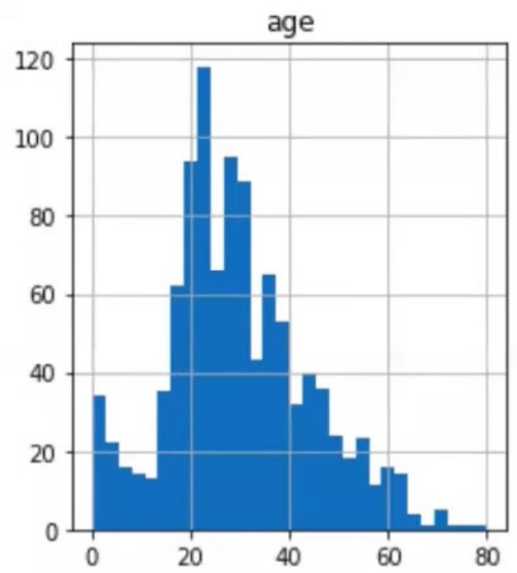
Add section

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 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 — ? — 10th percentile

0-16 16-20 20-22

10% 20% 30%

OneNote for Windows 10

Home Insert Draw View Help

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...

Add section

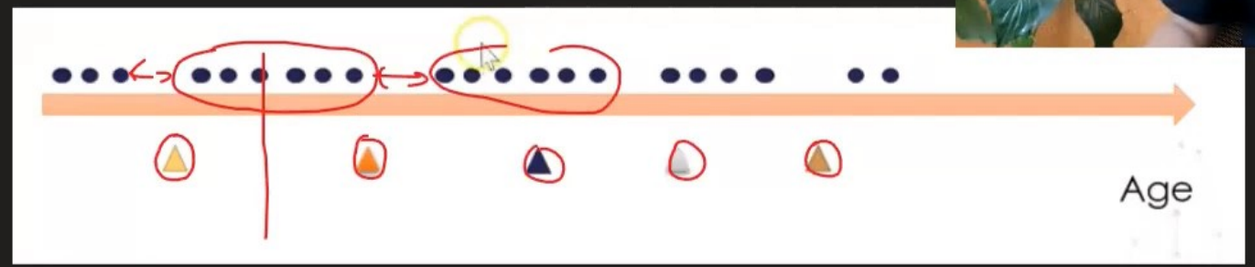
1. Encoding Numerical Features
2. Discretization
3. Types of Discretization
4. Equal Width/Uniform Binning
5. Equal Frequency/Quantile Binni...
6. KMeans Binning
7. Encoding the discretized variable
8. Example
9. Custom/Domain Based Binning
10. Binarization
11. Example

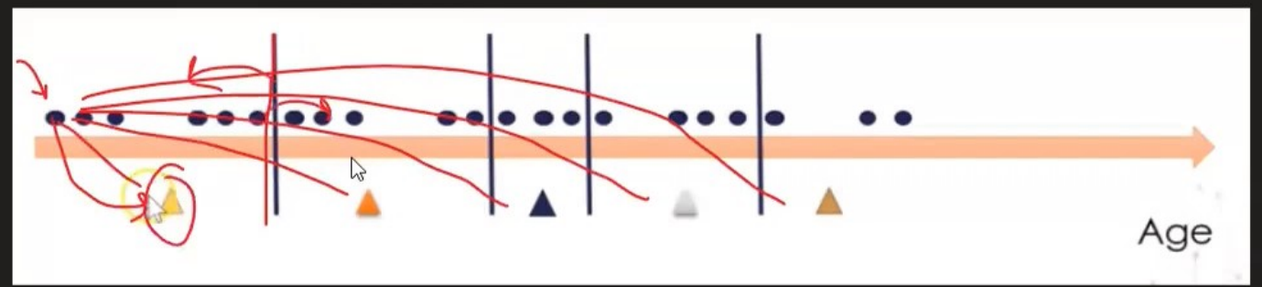
Add page

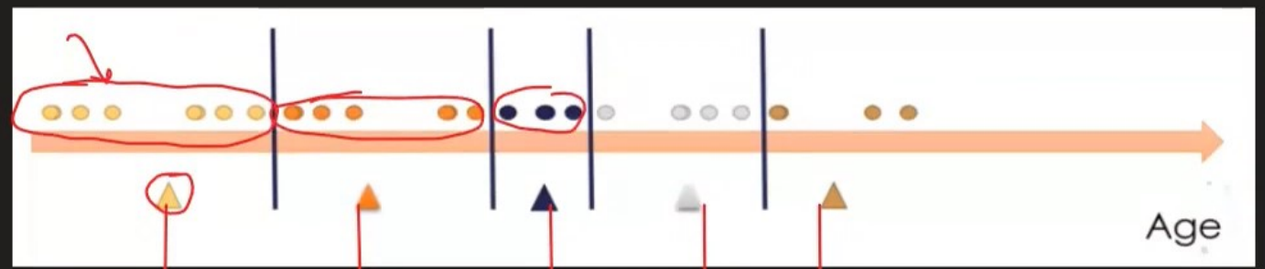
# 6. KMeans Binning

Monday, April 19, 2021 3:57 PM

centroid
interval-5





KMeans
↳ Clustering
2D
1 OD

OneNote for Windows 10

Home Insert Draw View Help

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...

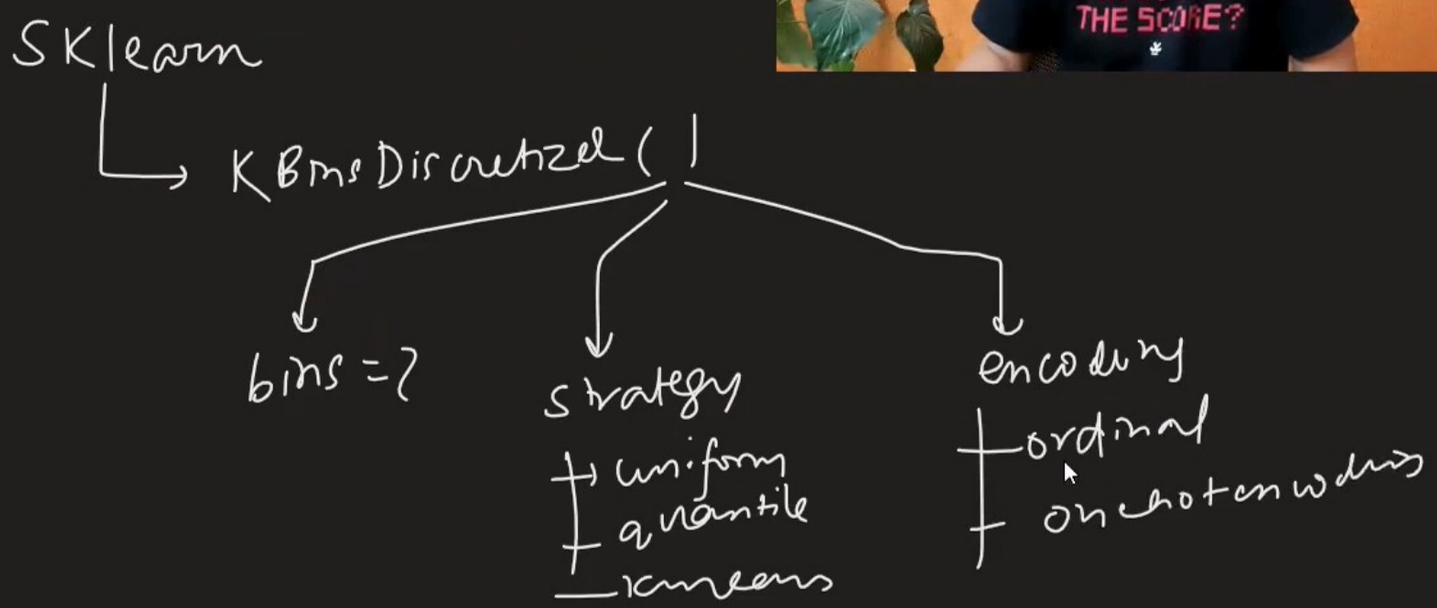
Add section

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 page

# 7. Encoding the discretized variable

Monday, April 19, 2021 3:58 PM




OneNote for Windows 10

Home Insert Draw View Help

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...

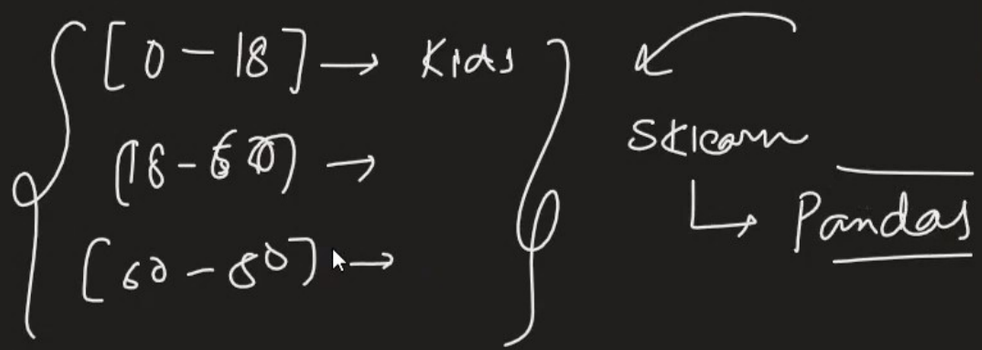
Add section

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 page

# 9. Custom/Domain Based Binning

Monday, April 19, 2021 3:58 PM




OneNote for Windows 10

Home Insert Draw View Help

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...

Add section

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 page

# 10. Binarization

Monday, April 19, 2021 4:17 PM

0,1

image

0-255 color

Annual income

62 <

0

1

>

127.5

0

1