CSE 310 L Data Warehousing and Mining

Date: 12-09-2023

1. Write a Python program to preprocess the date - decompose the date by mentioning three different columns for the parts of the date, such as Year, Month and Day.

Hint: Use pandas.to\_datetime

Example:

Input dataset:

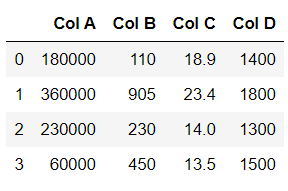
|  |  |
| --- | --- |
| Employee ID | Date of Join |
| 20005 | 05-01-2019 |
| 20006 | 10-01-2019 |
| 20007 | 01-02-2019 |
| 20008 | 05-02-2019 |
| 20009 | 01-03-2019 |
| 20010 | 02-03-2019 |
| 20011 | 02-04-2019 |
| 20012 | 05-04-2019 |

Output dataset:

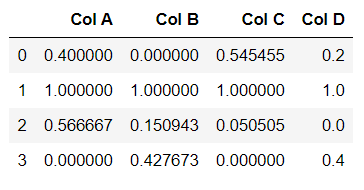
|  |  |  |  |
| --- | --- | --- | --- |
| Employee ID | Day | Month | Year |
| 20005 | 5 | 1 | 2019 |
| 20006 | 10 | 1 | 2019 |
| 20007 | 1 | 2 | 2019 |
| 20008 | 5 | 2 | 2019 |
| 20009 | 1 | 3 | 2019 |
| 20010 | 2 | 3 | 2019 |
| 20011 | 2 | 4 | 2019 |
| 20012 | 5 | 4 | 2019 |

1. Create the following data frame and do min-max normalization

Input:



Output:



Note: Min-max normalization formula

A black text on a white background

Description automatically generated