

Exploring Data with Pandas

CSC316 Machine Learning
Leon Tabak

08 February 2022

```
import numpy as np
import pandas as pd

numberOfColumns = 8
numberOfRows = 12

maxValue = 8

def makeRow(size):
    return [np.random.randint(0,maxValue) for n in range(size)]

def main():
    print( 'Guten Tag! ' )

    samples = [makeRow(numberOfColumns) for n in range(numberOfRows)]

    columnNames = [chr(n) for n in range(65, 65 + numberOfColumns)]
    rowNames = [chr(n) for n in range(97, 97 + numberOfRows)]

    data = pd.DataFrame( samples , columns = columnNames , index = rowNames )

    print( '\n' )

    print( data )

    print( '\nHead\n' )

    print( data.head() )

    print( '\nTail\n' )

    print( data.tail() )
```

```

print( '\nInfo\n' )

print( data.info() )

print( '\nValue_counts_in_column_A\n' )

print( data['A'].value_counts() )

print( '\nDescribe\n' )

print( data.describe() )

print( '\nMean_values_for_all_columns\n' )

print( data.describe().loc['mean'] )

print( '\nRows_in_which_value_in_column_A>3\n' )

print( data.loc[data['A'] > 3] )

print( f'\nJust_in_first_{numberOfColumns/2}\n' )

print( data.loc[:,
    [chr(n) for n in range(65, 65 + numberOfColumns//2)]] )

if __name__ == '__main__':
    main()

```