

## **Weekly Progress Report (Week 2)**

**Aniket Bhatia**

### **Task Completed**

The task completed during the week :

- Practiced numpy and pandas on a dataset
- Studied about the data science process
- Started working on the Smart City traffic pattern dataset

### **Challenges and Hurdles**

The major challenges that are encountered during the week :

- How to differentiate between train and test dataset
- How to overcome or solve the problem of null values in a dataset
- To understand how matplotlib works

### **Lessons Learned**

The knowledge gained during the week :

- Impact of Big Data on Business
- Data Scientist vs Data Analyst
- Facets of Data
- Data Science Process
- Methods of Numpy and Pandas

# Weekly Progress

## (Numpy and Pandas Practice)

```
[ ] array([[0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1 ],
          [0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, 0.2 ],
          [0.21, 0.22, 0.23, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.3 ],
          [0.31, 0.32, 0.33, 0.34, 0.35, 0.36, 0.37, 0.38, 0.39, 0.4 ],
          [0.41, 0.42, 0.43, 0.44, 0.45, 0.46, 0.47, 0.48, 0.49, 0.5 ],
          [0.51, 0.52, 0.53, 0.54, 0.55, 0.56, 0.57, 0.58, 0.59, 0.6 ],
          [0.61, 0.62, 0.63, 0.64, 0.65, 0.66, 0.67, 0.68, 0.69, 0.7 ],
          [0.71, 0.72, 0.73, 0.74, 0.75, 0.76, 0.77, 0.78, 0.79, 0.8 ],
          [0.81, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.88, 0.89, 0.9 ],
          [0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99, 1.  ]])
```

```
▶ np.linspace(0,1,20)
```

```
↳ array([0.          , 0.05263158, 0.10526316, 0.15789474, 0.21052632,
         0.26315789, 0.31578947, 0.36842105, 0.42105263, 0.47368421,
         0.52631579, 0.57894737, 0.63157895, 0.68421053, 0.73684211,
         0.78947368, 0.84210526, 0.89473684, 0.94736842, 1.        ])
```

```
[ ] mat = np.arange(1,26).reshape(5,5)
    mat
```

```
array([[ 1,  2,  3,  4,  5],
       [ 6,  7,  8,  9, 10],
       [11, 12, 13, 14, 15],
       [16, 17, 18, 19, 20],
       [21, 22, 23, 24, 25]])
```

```
[ ] mat[2:,1:]
```

```
array([[12, 13, 14, 15],
       [17, 18, 19, 20],
       [22, 23, 24, 25]])
```

```
[ ] myindex = ['USA','Canada','Mexico']
    mydata = [1776,1867,1821]
```

```
[ ] myser = pd.Series(mydata,myindex)
```

```
[ ] myser
```

```
USA      1776
Canada   1867
Mexico   1821
dtype: int64
```

```
▶ myser['USA'],myser[1]
```

```
↳ (1776, 1867)
```

```
[ ] ages = {'Sam':5,'Frank':10,'Spike':7}
```

```
[ ] pd.Series(ages)
```

```
Sam      5
Frank    10
Spike     7
dtype: int64
```

Conditional Filtering

```
[ ] df[df['Age'] > 32]
```

	Age	Gender	Education Level	Job Title	Years of Experience	Salary
2	45.0	Male	PhD	Senior Manager	15.0	150000.0
3	36.0	NaN	Bachelor's	Sales Associate	7.0	60000.0
4	52.0	Male	Master's	Director	20.0	200000.0
6	42.0	Female	Master's	Product Manager	12.0	120000.0
9	38.0	Male	PhD	Senior Scientist	10.0	110000.0
...	...	...	...	...	...	...
369	33.0	Male	Bachelor's	Junior Business Analyst	4.0	60000.0
370	35.0	Female	Bachelor's	Senior Marketing Analyst	8.0	85000.0
371	43.0	Male	Master's	Director of Operations	19.0	170000.0
373	34.0	Male	Bachelor's	Senior Operations Coordinator	7.0	90000.0
374	44.0	Female	PhD	Senior Business Analyst	15.0	150000.0

264 rows × 6 columns

& and | operators