

14 nth October Notes

#how to convert object into numerical

Checking Unique Values

```
final_data.Qty.unique()
```

Converting to Numeric

```
final_data['Qty']=pd.to_numeric(final_data['Qty'],errors='coerce')
```

```
final_data
```

```
#how to convert object into numerical
final_data.Qty.unique()
|
final_data['Qty']=pd.to_numeric(final_data['Qty'],errors='coerce')
final_data
```

	Date	Voucher	Party	Product	Qty	Rate	Gross	Disc	Voucher	Amount
0	1/4/2017	Sal:1	SOLANKI PLASTICS	DONA-VAI-9100	2.0	1,690.00	3,380.00	NaN		13,100.00
1	1/4/2017	Sal:1	SOLANKI PLASTICS	LITE FOAM(1200)	6.0	1,620.00	9,720.00	NaN		NaN
2	1/4/2017	Sal:2	SARNESWARA TRADERS	VISHNU CHOTA WINE	500.0	23	11,500.00	NaN		30,990.00
3	1/4/2017	Sal:2	SARNESWARA TRADERS	LITE FOAM(1200)	6.0	1,620.00	9,720.00	NaN		NaN
4	1/4/2017	Sal:2	SARNESWARA TRADERS	DONA-VAI-9100	5.0	1,690.00	8,450.00	NaN		NaN
...
111201	10/10/2019	Sal:4935	K.SRIHARI	13*16 WHITE RK	400.0	16	6,400.00	NaN		NaN
111202	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN		NaN
111203	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN		NaN
111204	NaN	Total	NaN	NaN	NaN	175,381.65	2,203,649.50	20,680.00		2,189,014.50
111205	NaN	Total	NaN	NaN	NaN	5,519,888.40	53,360,791.40	672,984.00		52,830,224.40

Note:

errors='coerce' converts invalid parsing to NaN.

Handling Null Values

Methods / Procedures to Handle Nulls:

1. **Ask your client** for correct or missing values
2. **Drop the null value records** → results in **data loss**
3. **Fill the null values** using **statistical methods**:

- Mean
- Median
- Mode

Checking Data Types

```
final_data.dtypes
```

using pivot table

```
pivot_table=pd.pivot_table(data=ins,values='expenses',index='region',columns=['smoker','children'],aggfunc='sum')
```

```
pivot_table
```

```
#using pivot table
pivot_table=pd.pivot_table(data=ins,values='expenses',index='region',columns=['smoker','children'],aggfunc='sum')
pivot_table
```

	smoker		yes										
	no												
children	0	1	2	3	4	5	0	1	2	3	4	5	
region													
northeast	976747.39	544403.10	490110.44	221947.50	101396.36	20936.92	732342.66	711482.80	204262.33	340039.14	NaN	NaN	
northwest	814816.76	518805.68	547296.76	348081.14	46609.63	8965.80	680000.21	238233.23	341347.99	470082.25	21472.48	NaN	
southeast	826029.55	585202.76	360936.28	287678.69	72255.12	60692.65	1420619.78	715066.23	677142.84	358065.90	NaN	NaN	
southwest	876219.52	535306.23	357878.89	276805.98	46408.47	48530.01	771294.22	276399.64	638679.83	108084.41	58124.36	19023.26	

Checking Region Counts

```
ins.region.value_counts()
```

```
ins.region.value_counts()
```

	count
region	
southeast	364
southwest	325
northwest	325
northeast	324

```
dtype: int64
```

Numpy

#its a python library used for numerical computations,especially with large multi-dimensional

#arrays and matrices.

#It also provide many mathematical,logical,and statiststical operations

Installation and Import

`!pip install numpy`

`import numpy as np`

`np.__version`

Creating Arrays

`arr1=np.array([1,3,2.4,5,7.8])`

`arr1`

List vs Array

`l1=[1,2,3,4,5]`

`l2=[4,5,6,7,8]`

`l1+l2 #concatenation in lists`

`arr1=np.array([1,2,3,4,5])`

`arr2=np.array([4,5,6,7,8])`

`arr1+arr2 #element-wise addition in arrays`