```
Longest Distance: 1000 m
Longest Duration: 1000 sec
Longest Speed: 10.0 m/s
Longest Altitude: 1000 m
Longest Calories: 10.0 kcal
```

```
# Shape and info from dataset
```

```
df.shape
```

```
(1000, 15)
```

```
# Examine variables and their values
```

```
df.head
```

	id	WhatWeAre	Walking	LastUsage	Lasting	Hour	Day	LastDays	LastFunc	Calory	Positive	Profit	Tech	Ob
0	1	0	0	1	1000	0	Sun	Sat	Sig	10	Office	1	Sat	Tech
1	2	1	10	0	900	1000	Sun	Sat	Sig	10	Office	0	Sat	Tech
2	3	2	20	1	800	1100	Sun	Sat	Sig	10	Office	1	Sat	Tech
3	4	3	30	0	600	1200	Sun	Sat	Sig	10	Office	0	Sat	Tech
4	5	4	40	1	500	1300	Sun	Sat	Sig	10	Office	1	Sat	Tech
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
995	996	0	0	0	1000	1011	Sun	Sat	Sig	10	Office	0	Sat	Tech
996	997	1	10	1	900	1012	Sun	Sat	Sig	10	Office	1	Sat	Tech
997	998	2	20	0	800	1013	Sun	Sat	Sig	10	Office	0	Sat	Tech
998	999	3	30	1	600	1014	Sun	Sat	Sig	10	Office	1	Sat	Tech
999	1000	4	40	0	500	1015	Sun	Sat	Sig	10	Office	0	Sat	Tech

1000 rows × 15 columns

```
# Checking the types of the dataset
```

```
df.dtype
```

```
id            int64
dtype: object
```

```
# Identify variables with nulls values
```

```
df.isnull.sum
```

```
id            0
WhatWeAre     0
Walking       0
LastUsage     0
Lasting       0
Hour          0
Day           0
LastDays      0
LastFunc      0
Calory        0
Positive      0
Profit        0
Tech          0
Obsession     0
dtype: object
```

```
# Identify variables with nulls values
```

```
df.isnull.sum/df.shape[0]*100
```

```
id            100.0
WhatWeAre     100.0
Walking       100.0
LastUsage     100.0
Lasting       100.0
Hour          100.0
Day           100.0
LastDays      100.0
LastFunc      100.0
Calory        100.0
Positive      100.0
Profit        100.0
Tech          100.0
Obsession     100.0
dtype: object
```

```
# Identify variables with nulls values
```

```
df.isnull.sum()[df.isnull.sum()!=0]
```

```
id            1000
WhatWeAre     1000
Walking       1000
LastUsage     1000
Lasting       1000
Hour          1000
Day           1000
LastDays      1000
LastFunc      1000
Calory        1000
Positive      1000
Profit        1000
Tech          1000
Obsession     1000
dtype: object
```

```
# Identify variables with nulls values
```

```
df.isnull.sum()[df.isnull.sum()!=0]/df.shape[0]*100
```

```
id            100.0
WhatWeAre     100.0
Walking       100.0
LastUsage     100.0
Lasting       100.0
Hour          100.0
Day           100.0
LastDays      100.0
LastFunc      100.0
Calory        100.0
Positive      100.0
Profit        100.0
Tech          100.0
Obsession     100.0
dtype: object
```

```
# Identify variables with nulls values
```

```
df.isnull.sum()[df.isnull.sum()!=0].index.tolist
```

```
['id',
 'WhatWeAre',
 'Walking',
 'LastUsage',
 'Lasting',
 'Hour',
 'Day',
 'LastDays',
 'LastFunc',
 'Calory',
 'Positive',
 'Profit',
 'Tech',
 'Obsession']
```

```
# Identify variables with nulls values
```

```
df.isnull.sum()[df.isnull.sum()!=0].index.tolist()
```

```
['id',
 'WhatWeAre',
 'Walking',
 'LastUsage',
 'Lasting',
 'Hour',
 'Day',
 'LastDays',
 'LastFunc',
 'Calory',
 'Positive',
 'Profit',
 'Tech',
 'Obsession']
```

```
# Identify variables with nulls values
```

```
df.isnull.sum()[df.isnull.sum()!=0].index.tolist()/df.shape[0]*100
```

	W	Whatever	Lottery	Everything	Everything	Fashion	Authentication	Machinery	Accuracy	Accuracy	Accuracy	Woodland
1	1	0.973	1	1	1	1.000	1.000	1.000	1.000	1	1	1
2	2	0.973	1	1	1	1.000	1.000	1.000	1.000	1	1	1
3	3	0.973	1	1	1	1.000	1.000	1.000	1.000	1	1	1

1970	12	1	0000	0	0	1966	1966	0	0	0	0
1970	12	1	0100	0	0	1966	1966	0	0	0	0
1970	12	1	0200	0	0	1966	1966	0	0	0	0
1970	12	1

1000	1010	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1020	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1030	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1040	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1050	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1060	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1070	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1080	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1090	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1100	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1110	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1120	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1130	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1140	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1150	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1160	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1170	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1180	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1190	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1200	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1210	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1220	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1230	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1240	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1250	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1260	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1270	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1280	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1290	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1300	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1310	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1320	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1330	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1340	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1350	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1360	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1370	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1380	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1390	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1400	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1410	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1420	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1430	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1440	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1450	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1460	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1470	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1480	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1490	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1500	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1510	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1520	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1530	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1540	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1550	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1560	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1570	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1580	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1590	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1600	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1610	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1620	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1630	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1640	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1650	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1660	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1670	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1680	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1690	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1700	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1710	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1720	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1730	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1740	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1750	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1760	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1770	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1780	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1790	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1800	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1810	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1820	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1830	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1840	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1850	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1860	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1870	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1880	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1890	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1900	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1910	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1920	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1930	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1940	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1950	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1960	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1970	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1980	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	1990	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	
1000	2000	0.000	0	0	1000	1000.0	1000.0	0.0	0	0	

1000 rows × 10 columns

```
In [14]: df.info()
```

```
In [15]: df.describe()
```

```
In [16]: # Checking correlation of column with other numeric values
```

```
In [17]: sns.pairplot(df,vars=['Open'])
```

```
In [18]: sns.pairplot(df,vars=['Open','Close'])
```

```
In [19]: # Checking for null values
```

```
In [20]: df.isnull().sum()
```

```
In [21]: # Plot showing some information, relationship from both columns
```

```
In [22]: sns.jointplot(x='Close',y='Open',data=df,kind='reg')
```

```
In [23]: sns.jointplot(x='Close',y='Open',data=df,kind='hex')
```

Identify the skewness and distribution

```
In [24]: numeric_cols.hist(figsize=(12,12))
```

```
Out[24]: array([[<AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>],
                [<AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>],
                [<AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>],
                [<AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>],
                [<AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>],
                [<AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>],
                [<AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>],
                [<AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>],
                [<AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>],
                [<AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>],
                [<AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>],
                [<AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>,
                 <AxisSubplot:>]]
```

```
Correlation Matrix
```

```
[1] 1.0 0.8 0.6 0.4 0.2 0.0 -0.2 -0.4 -0.6 -0.8 -1.0
Warning message:
Ignoring unknown parameters: "Transaction", "Customer", "Product", "Quantity", "Price", "Discount", "Tax", "Total", "Payment", "Delivery", "Rating", "Feedback", "Promotion", "Publication", "Publication_Year", "Publication_Decade", "Pages", "Genre", "Publication_Type", "Language", "Publication_Country", "Publication_Region", "Author", "Author_Gender", "Author_Nationality", "Author_Birth_Year", "Author_Birth_Decade", "Author_Age", "Author_Type", "Author_Region", "Author_Publication_Count", "Author_Publication_Frequency", "Author_Publication_Span", "Author_Publication_Rate", "Author_Publication_Recency", "Author_Publication_Consistency", "Author_Publication_Diversity", "Author_Publication_Impact", "Author_Publication_Quality", "Author_Publication_Trend", "Author_Publication_Seasonality", "Author_Publication_Cycle", "Author_Publication_Pattern", "Author_Publication_Anomaly", "Author_Publication_Outlier", "Author_Publication_Cluster", "Author_Publication_Segment", "Author_Publication_Profile", "Author_Publication_Score", "Author_Publication_Rank", "Author_Publication_Percentile", "Author_Publication_Quartile", "Author_Publication_Decile", "Author_Publication_Ratio", "Author_Publication_Index", "Author_Publication_Weight", "Author_Publication_Value
```

Ecuador
Dominica
Cuba
Chile
Bolivia
Belize
Argentina
Antigua
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025
0.0030
0.0035
0.0040
0.0045
0.0050
0.0055
0.0060
0.0065
0.0070
0.0075
0.0080
0.0085
0.0090
0.0095
0.0100
0.0105
0.0110
0.0115
0.0120
0.0125
0.0130
0.0135
0.0140
0.0145
0.0150
0.0155
0.0160
0.0165
0.0170
0.0175
0.0180
0.0185
0.0190
0.0195

Count
Ms
Mr
Mrs
Miss
Title

Count
Male
Female
Sex

Count
Cha
Q
S
Embarked

Count
1
2
3
Pclass

Count
Alone
Not Alone
Alone

Count
Child
Teenager
Young Adult
Adult
Senior
Age Group

Count
Low
Medium
High
Very High
Fare Group

Count
Cabin
No Cabin
Has Cabin

Count
Unknown
Cabin A
Cabin B
Cabin C
Cabin D
Cabin E
Cabin F
Cabin G
Cabin T
Cabin Group

Count
Unknown
Mr
Mrs
Miss
Master
Rare
Title Group

Count
Unknown
Mr
Mrs
Miss
Master
Rare
Title Group

Count
Unknown
Mr
Mrs
Miss
Master
Rare
Title Group

Count
Unknown
Mr
Mrs
Miss
Master
Rare
Title Group

Count
Unknown
Mr
Mrs
Miss
Master
Rare
Title Group

Count
Unknown
Mr
Mrs
Miss
Master
Rare
Title Group

Count
Unknown
Mr
Mrs
Miss
Master
Rare
Title Group

Count
Unknown
Mr
Mrs
Miss
Master
Rare
Title Group

Count
Unknown
Mr
Mrs
Miss
Master
Rare
Title Group

Count
Unknown
Mr
Mrs
Miss
Master
Rare
Title Group

Count
Unknown
Mr
Mrs
Miss
Master
Rare
Title Group

Count
Unknown
Mr
Mrs
Miss
Master
Rare
Title Group

Count
Unknown
Mr
Mrs
Miss
Master
Rare
Title Group

Count
Unknown
Mr
Mrs
Miss
Master
Rare
Title Group

Count
Unknown
Mr
Mrs
Miss
Master
Rare
Title Group